Virtual offices

Peter Robinson
University of Cambridge Computer Laboratory[†]

ABSTRACT

A joint project between the University of Cambridge Computer Laboratory and the Rank Xerox Research Centre in Cambridge has been looking at ways of using digitised video from television cameras in user interfaces for computer systems.

The *DigitalDesk* is built around an ordinary physical desk and can be used as such, but it has extra capabilities. A video camera is mounted above the desk, pointing down at the work surface. This camera's output is fed through a system that can detect where the user is pointing and it can read documents that are placed on the desk. A computer-driven projector is also mounted above the desk, allowing the system to project electronic objects onto the work surface and onto real paper documents.

Another approach is used in *BrightBoard*. A video camera is pointed at an ordinary whiteboard and its output fed into a computer. The whiteboard thus becomes an alternative means of controlling the computer.

These systems show how computers can be built into everyday objects with simple user interfaces that do not require expert knowledge to operate. As such they exemplify a new approach to human-computer interaction where the computer is brought into our offices instead of squeezing an office uncomfortably into a computer screen.

INTRODUCTION

In the 1970s, scientists at the Xerox Corporation's Palo Alto Research Center developed the *desktop metaphor* which made computers easy to use by making them look and act like ordinary physical desks. Electronic documents could be manipulated like paper documents, but the computer added powerful new facilities. This led some to predict that the *paperless office* would dominate within a few years. The trouble is that people like paper. It's portable, tactile and easier to read than a screen; today computers are used to generate far more paper than they replace.

The same can be said of many other kinds of office paraphernalia. We are flooded by electronic diaries, electronic whiteboards, electronic mail and so on. These all offer valuable improvements by comparison with their physical originals, but often at the expense of general convenience and ease of use. An electronic diary is bigger than a pocket diary, writing on an electronic whiteboard with a mouse is more difficult than using a pen, it is hard to jot notes for a reply in the margin of a piece of electronic mail.

An alternative approach is to add computational properties to the conventional office environment. This is not *virtual reality* where the user is immersed in a totally synthetic, computer-generated environment, often donning a special headset and even clothes; this is *augmented reality* where the computers augment the everyday, real world. This requires the computer to monitor activities and to deliver its contribution as unobtrusively as possible, suggesting the use of video and, to a lesser extent, sound for input and output. Of course, this merely reflects normal office practice. The virtual office should be a seamless extension of the normal physical office.

[†] *Full address:* University of Cambridge, Computer Laboratory, New Museums Site, Pembroke Street, Cambridge, England CB2 3QG. *Telephone:* +44 1223 334637. *Facsimile:* +44 1223 334678. *E-mail:* pr@cl.cam.ac.uk *URL:* http://www.cl.cam.ac.uk/users/pr/

Recent developments in computer hardware are greatly reducing the cost of attaching television cameras to computers. They have moved from being an expensive peripheral for specialists to a level comparable today with a monitor and developments in technology will soon make the cost similar to that of a mouse. This raises the question of what new techniques will be appropriate when every computer routinely includes video input, possibly from several cameras.

This paper describes a number of experiments that have been undertaken in the University of Cambridge Computer Laboratory and at the Rank Xerox Research Centre in Cambridge the use of video in computer augmented environments. Two prototype systems are discussed in some detail, both using video input but differing in the amount of feedback presented to the user, and a number of variants of these are mentioned.

Approaches to the three principal technical problems are presented:

- **thresholding** the image from a camera to give an easily-processed binary image,
- **registration** of the camera and any projected images with the physical world, and
- **user interfaces** appropriate for the new technology.

Finally, some topics of continuing research are identified.

DIGITALDESK

The *DigitalDesk* [Wellner 1994] is based around an ordinary physical desk and can be used as such, but it has extra capabilities. A video camera is mounted above the desk, pointing down at the work surface. This camera's output is fed through a system that can detect where the user is pointing and it can read documents that are placed on the desk. A computer-driven projector is also mounted above the desk, allowing the system to project electronic objects onto the work surface and onto real paper documents (something, incidentally, that can not be done with flat display panels or rear-projection).

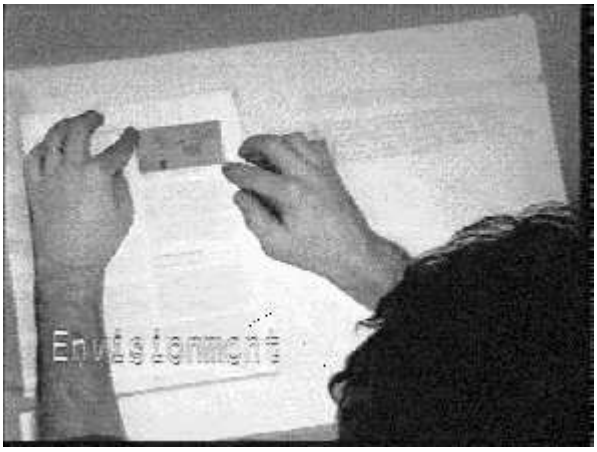
In one sense, this just gives the effect of a computer screen on the desk-top (instead of the more common desk-top on a computer screen). Indeed, this emulation can be made precise. In the first implementation [Wellner 1991], image processing software was used to identify the outline of a hand and its pointing finger in the image, and its location was delivered to the window system as if it were a conventional pointing device such as a mouse. A microphone fixed under the desk detected a sharp noise such as a finger tapping and delivered it to the computer as if a mouse button had been pushed. Thus pointing and tapping on the desk with a finger modelled pointing and clicking with a mouse. Standard application programs, such as a four-function calculator, could be driven simply by gesticulating with the hands.

The prototype DigitalDesk calculator

As it stands, this seems to offer little advantage compared with a conventional calculator. The difference becomes apparent when we recall that a common use of calculators involves the transcription of figures from a sheet of paper into the calculator for further processing. With the

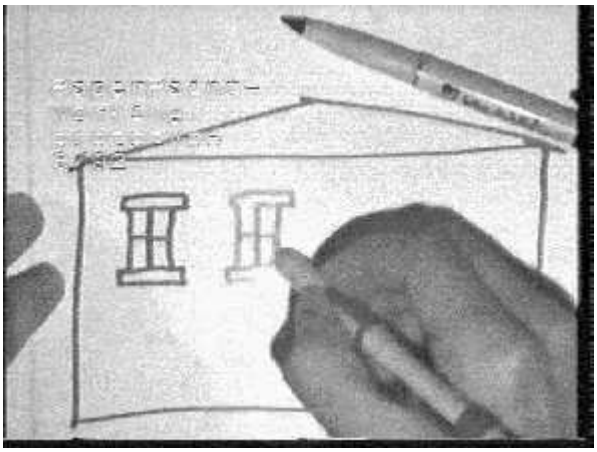
DigitalDesk no re-keying is necessary. A piece of paper containing figures can be placed on the desk, a number selected by pointing with the finger and the value copied into the calculator. This works by capturing the image of the figures, passing it through optical character recognition software and presenting the result as the current selection to the computer's window system. This can be thought of as copy and paste from a physical document into an electronic one.

A more radical application would allow the user to pick up a pencil and start writing on a sheet of paper, with the computer identifying the activity and interpreting it as a specialised data entry operation. For example, it might create a new text file, spreadsheet or drawing as appropriate to hold the actual data. The research is still working towards this goal.



Simulated word processor, spreadsheet and drawing program [Wellner 1992]

One step along the way is *PaperPaint,* is a simple painting program that uses the DigitalDesk to extend its facilities. Pictures are drawn with a pen on paper in the usual way but the overhead camera can capture parts of the image so that they can be copied and superimposed elsewhere in the drawing by the overhead projector. This allows parts of the drawing to be copied and added to the image. Normal painting program operation can be applied to the projected copies, so they can be moved around, but they are also properly part of the image and are captured if a further selection is made. In particular, the combined image can be captured and passed to a printer to deliver a hard-copy version of the drawing.

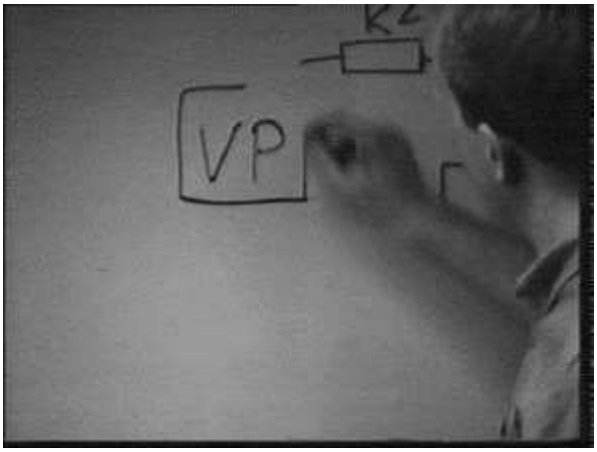


Working with PaperPaint

In practice, it turned out that designers used this facility in an unexpected way. Fragments of drawings were prepared on separate pieces of paper and these were then placed over the image being prepared before being selected, copied and pasted into place. This gives an effect rather like dry-transfer lettering with the added advantage that the applied image can be moved subsequently. The user interface is also particularly attractive - there are no complicated commands to rotate an applied image. The fragment is simply positioned where it is wanted before copying, exploiting the natural physical skills of the user. Similarly, a library of clip art can be kept on paper, making it much more accessible to those less versed in technology.

BRIGHTBOARD

The second major experiment in this work is *BrightBoard* [Stafford-Fraser 1995]. This dispenses with the projection system of the DigitalDesk and concentrates on using video input to control a computer. The underlying model is to take another standard piece of equipment - a whiteboard or a flip-chart of even a sheet of paper - and point a camera at it to create some of the properties of electronic documents. The images on the board can be saved, printed, faxed and e-mailed simply by walking up to the board and writing the corresponding command. BrightBoard can also operate as a more general input device for any computer-controlled system such as air-conditioning or video recording.



Writing a command with BrightBoard

The challenge is to separate commands from the more general writing on the whiteboard. The user wants to write with an ordinary pen on the whiteboard, so that no special electronic pen or additional sensor installation is required. This means that commands have to be recognised from a noisy signal by image processing. This is handled in three phases:

- identify marks from an alphabet of about 20 symbols,
- identify a group of marks having a particular spatial relationship to each other, and
- execute a command associated with the particular group of commands.

Each of these phases is controlled by configuration settings appropriate for an individual user. Typical symbols are letters, digits, boxes and corner marks to delimit an area on the board. A typical group for a command would be to write mnemonic letters inside a box. This is reminiscent of selecting a command from a menu on a conventional computer system. It is also unlikely to appear as part of the other writing on the board, avoiding false triggering of commands.

Three further issues arise in the user interface to this system. The first is that commands to capture the image on the whiteboard and, for example, print it yield a printed copy that includes the written print command and possibly other extraneous information. The user needs to select an area of interest on the board and only have the commands applied to that area. This is achieved by writing special symbols at the corners of the selected area and having the software only apply operations to the corresponding subset of the image.

The second problem is the fact that the user is often standing between the camera and the whiteboard and so obscures the camera's view. Interestingly, this was not a problem with the DigitalDesk because its users tended to keep their hands clear of the area of interest so that they could see it themselves. The larger scale of operation of BrightBoard introduces the new problem. The solution adopted is to make the system sensitive to movement in front of the whiteboard. Differences are calculated between successive images at a fairly coarse resolution, significant differences indicate movement and small differences indicate stability. The system waits for a period of movement, indicating a user at work, followed by a period of stability, indicating the conclusion of the activity, after which a full resolution image is captured and processed. Users build up a mental model of using

the whiteboard and then standing back to let the system have a look when they want it to take some action, which is rather like making a presentation to a human audience.

Finally, there is the question of giving confirmation to the user that an action has been triggered and avoiding multiple responses to a single command. In the absence of a projector, this has been resolved by using speech synthesis in the controlling computer. Having written a command, the user stands back to let the computer analyse the contents of the whiteboard and invoke the corresponding commands. Each command includes a simple confirmation to be spoken by the synthesiser. Having heard this, the user can return to the board and erase the command.

ASSOCIATED WORK

The ideas embodied in the DigitalDesk and BrightBoard have been applied in a number of different other experimental projects.

Marcel

Marcel [Newman & Wellner 1992] is a desktop translation assistant. A book or document in French is placed on the desk and read in the ordinary way. When the reader encounters an unfamiliar word, it is selected with the finger and possible translations are projected onto the desktop next to the book.

In principle, the camera should capture the image of the French word and pass it through an optical character recognition system to produce the text to be looked up in a dictionary. However, the resolution of the overhead camera is not really good enough to support this. One approach is to use a second camera zoomed tightly in on a small part of the desk giving sufficient resolution for OCR. Marcel uses a different approach. It is assumed that all the text to be translated is available as high resolution digitised images, either because the book has been scanned into the computer beforehand or because it was generated by computer in the first place. Each page can then be recognised by calculating a signature encoding the general shape of the page as a function of line lengths, paragraph heights and word breaks. The position of the word for translation can then be mapped back onto the high resolution image and the text extracted.

Digital Drawing Board

The *Digital Drawing Board* [Carter 1993] is a version of the DigitalDesk for use in Computer Aided Design and again aims to supplement conventional working practices rather than to replace them. The system embeds a large (A1 size) digitising tablet in an architect's drawing board that also has a television camera and projector directed at it. The designer works in the usual way with pencil, paper and other conventional tools, but any image on the drawing board can be captured by the camera, processed by a computer and projected back into a new window on the drawing board

A prototype application has been built that allows a user to sketch the cross section of a solid and a texture to be mapped onto it. These are captured by the system, the solid of revolution calculated and a perspective view projected with the texture applied to the surface. This is appropriate for the early stages of architectural design where many different arrangements of volumes and textures are being explored.

EuroCODE

EuroCODE [RXRC 1993] is the European CSCW (Computer Supported Co-operative Work) Open Development Environment, an ESPRIT-III project investigating support for geographically separated collaborators. Part of the project is a multimedia computer system integrating text, video and audio, together with other media types. The system uses DigitalDesk technology to include paper among the integrated media.

This is being used by contractors involved in the Great Belt bridge construction project in Denmark, allowing engineering drawings to be shared by workers at opposite sides of the waterway. The idea is to preserve the conventional, tactile benefits of paper documents while adding new benefits in the

form of shared remote access. Engineers work with paper drawings on an ordinary desk, but annotations in the form of sketches or sound and video recordings can be attached to the relevant location on the document. These can then be recovered by pointing at an area of interest on the current drawing.

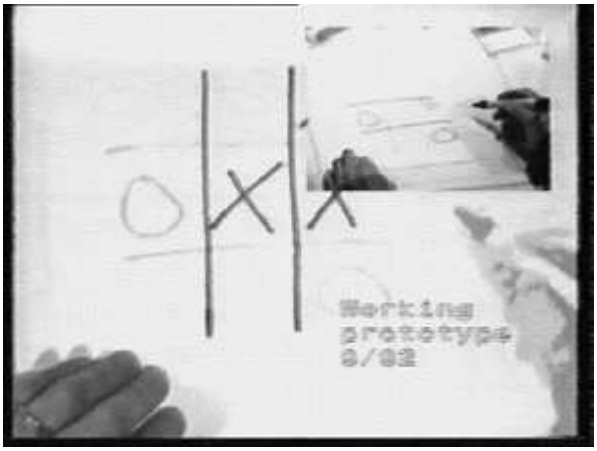
Mosaic

Mosaic [Mackay *et al* 1993] is a tool for manipulating story-boards for video production. A story-board consists of a sequence of drawings showing key frames in a collection of video clips to be edited together. With Mosaic, these are created by drawing on paper or printing still pictures from the clips and then assembling these on a DigitalDesk. These key frames are then captured by the overhead camera and the film edited into the corresponding order.

This can then be viewed directly by composing the clips together from a write-once video disc (although digital video processing would be equally effective) and the results projected back onto the desk-top. Each time the story-board is edited, the effect on the complete production can be checked by replaying the clips in the new order.

DoubleDigitalDesk

The *Double DigitalDesk* [Freeman 1994] consists of two DigitalDesks connected by a computer data network. Each desk repeatedly grabs its image and passes it to the other desk for display so that both users see the other's desk merged with their own. Some care has to be taken with transforming the images to ensure that registration is maintained and in controlling the level of feedback.



The Double DigitalDesk with the remote user's desk inset

Similar experiments in the past [Krueger 1985] used analogue video transmission, but the mechanical problems of adjusting cameras and projectors to ensure registration made them impracticable. There was also no opportunity for additional computer assistance. The digital version solves the registration problem by extending the transformations already used to deal with the camera and projector and generally offers a more interesting vehicle for computer-supported co-operative work.

TECHNICAL ISSUES

Several interesting technical problems have had to be solved in the construction of the two principal prototypes. The resulting infrastructure has been used in the further associated work and is being used in current experiments. This has been made particularly easy by the use of Modula-3 [Nelson] as the implementation language and its support for the re-use of library code has been especially valuable.

Thresholding

The first processing problem is to extract a usable image from the camera over the desk or pointed at the whiteboard. The lighting is uneven, varying both spatially across the desk-top and with time as the environment changes. Moreover, the background image itself changes as different pieces of paper and other objects are placed on the desk, so it is not possible simply to subtract out the background. Significant information has to be separated from this noise. Conventional image processing techniques are available, but they are too expensive in terms of computing power to be useful for interactive applications.

The solution is to use an adaptive thresholding algorithm to extract a binary image from the scanned grey levels. This sweeps through the image building up a running average of the intensity levels encountered as a weighted combination of the preceding pixels. The direction of travel across the image alternates on successive scan lines and the values also averaged down the picture. This running average is then used as a threshold against which each pixel can be compared and quantised to a black or white value.

The net effect is that a black and white image can be extracted from the grey level image in a single pass over the pixel array and the result is relatively unaffected by uneven illumination.

Registration

The second area is the registration of the physical desk-top with the digitised and projected images. Occasionally, a digitising tablet is used instead of trying to follow the user's finger in which case this too had to be registered with the other co-ordinate systems. This was further compounded in the Double DigitalDesk by having two complete sets of co-ordinate systems which must all be registered with each other.

Unfortunately mechanical sloppiness in the system means that it can not be assumed that any of these are accurately aligned. The axes of the cameras and projectors may not be perpendicular to the desk-top, giving rise to keystone distortion, and may be rotated with respect to each other. This is further compounded distortions introduced by the various lens systems. The net effect is that simple transformations using only translation and scaling are not sufficient.

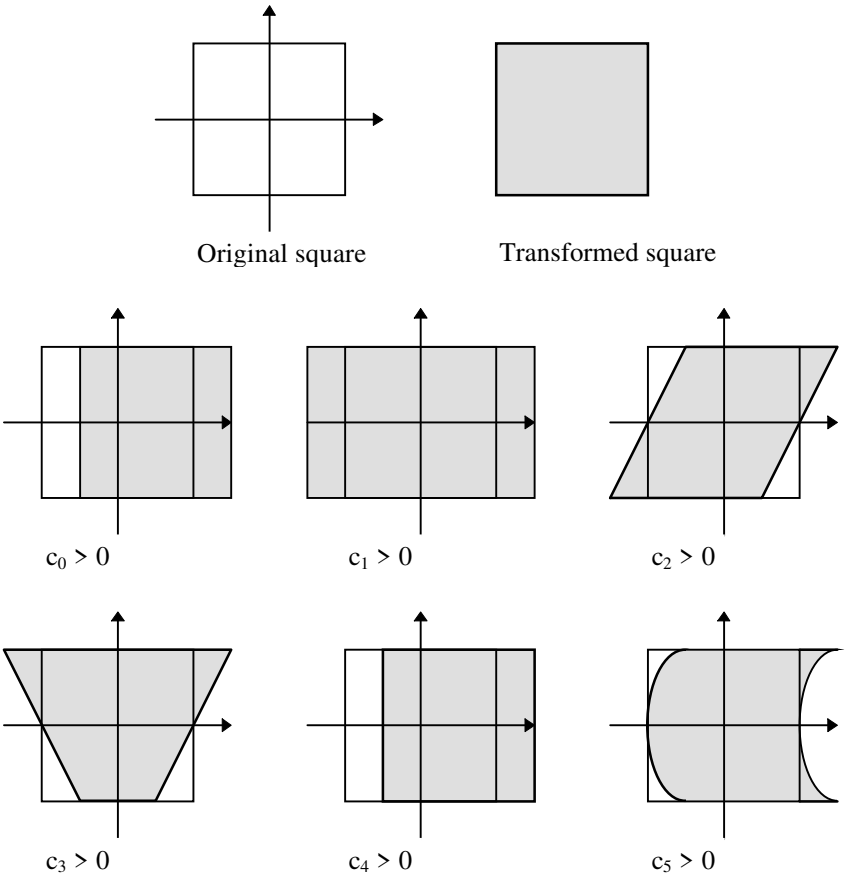
The next step in complexity would be to introduce quadratic terms. Consider the transformation calculating *x'* in terms of *x* and *y*:

$$x' = c_0 + c_1 x + c_2 y + c_3 xy + c_4 x^2 + c_5 y^2$$

The six coefficients each correspond to a different type of distortion encountered in projection:

- c_0 translation
- c_1 uniform scaling
- c_2 shearing
- c_3 keystoning
- c_4 non-uniform scaling
- c_5 bending

These can be pictured as follows [Wellner 1994]:



Of these, the first four are most significant, remembering that a rotation can be approximated by combining shearing in both the *x* and *y* directions.

A similar equation expresses *y'* as a linear combination of *x*, *y* and *xy*. This gives a set of eight coefficients to be calculated from calibration data which is possible using four sample registration points. For the camera, this is performed automatically by projecting four marks at the corners of the field of view, searching for them and noting their positions. For the digitising tablet, the user has to point the stylus at the registration marks when they are projected.

User interface

A user interface built from video capture rather than physical tools such as a mouse or keyboard allows a much richer vocabulary for interaction. Context-sensitive menus can be projected onto the DigitalDesk, but this is often unnecessary as instructions can be conveyed by manipulation of the documents themselves. BrightBoard is more challenging because commands have to be separated from the ordinary text written on the board.

The three phases involved were mentioned earlier. After thresholding the image to black and white, connected components of black writing on a white background are identified using a simple flood-fill algorithm. A number of metrics of the component are then calculated. These are chosen to be independent of scale and of the thickness of the pen used. For example, one value might be the ratio of the number of black pixels with white above them to the number of black pixels with white to the right of them, which gives and indication of the ratio of horizontal to vertical lines in the symbol. Others are based on various moments about the centre of gravity of the component.

The set of metrics defines a point in a multi-dimensional space. This is compared with the positions in the same space of instances of prototype symbols analysed during a preliminary training session. The scales of the different axes can not be compared, so all values are reduced by dividing by the standard deviation of values found in that dimension amongst the prototypes. This is effectively Mahalanobis distance rather than Euclidean distance. Also, rather than identifying a symbol with the nearest prototype, the dominant symbol in a small collection of neighbouring prototypes is chosen.

Having identified a symbol in the system's vocabulary, the next step is to see if it constitutes a command or should be ignored as part of the more general writing on the whiteboard. It would be rash simply to act on the presence of a single symbol, the system needs to recognise a specific configuration of symbols associated with an action to be invoked. Moreover, the configurations used are going to be different for different users.

The solution that has been adopted is for the first phase to record the symbols that it has detected in a Prolog database as rules associating the identity of the particular symbol with its position and with the prototype that it matched. The user can then define goals combining these primitives. For example, a print command might be defined as follows:

```
doprint :-
   issym(X, p),
   issym(Y, checkbox),
   inside(X, Y),
   /+ (inside(Z, Y), Z \= X).
```

This can be read as, "There is a print command if the letter 'P' appears inside a box and nothing else is inside the box." In fact the Prolog database records both the current and the previous state of the whiteboard, so the rules can be elaborated to deter multiple execution of a single command on successive analyses.

Finally, a further configuration file links predicates such as `doprint` to Unix commands that will actually do the work. These can use images from the whiteboard or a selected subset of it as arguments.

CURRENT WORK

Work is continuing on developing the general infrastructure for video user interfaces, investigating new methods for human-computer interaction, and applying and evaluating them in the specific discipline of electronic publishing. The goal is to combine electronic and printed documents to give a richer presentation than that afforded by either separate medium. We hope to prove the concept of combined documents and to pave the way for subsequent development of mixed-media publications.

Electronic, multi-media publishing is becoming established as an alternative to conventional publishing on paper. CD-ROM versions of reference books and fiction can augment their conventional counterparts in a number of ways:

- They can offer elaborate indexing, glossaries and cross-referencing.
- They can allow non-linear progression through the text. This can be used to especially good effect in educational material, where each student can be guided through in a way tailored to his or her particular needs, determined by questions and answers.
- Sound and moving images can be added in a multi-media presentation.
- Sections can easily be copied into new documents.

However, screen-based documents have a number of disadvantages:

- People find screens harder to read than paper.
- Electronic bookmarks are far less convenient that simply keeping a finger in the appropriate page.
- Adding personal notes to electronic documents is at best difficult and often impossible.
- Writing, editing and proof-reading a non-linear, multi-media document is still a specialised and difficult task.

The solution is to publish material as an ordinary, printed document that can be read in the normal way, enjoying the usual benefits of readability, accessibility and portability. However, when observed by a camera connected to a computer, it will acquire the properties of an electronic document.

This entails work on the underlying technology and on new applications to prepare and present mixed-media publications. Three key areas of research are involved. The first is video resolution. New projectors and cameras are becoming available that allow laser printer resolution for desktop image projection and capture. In the mean time these can be approximated by combining two or more low resolution devices.

This leads to the question of character and image: the printed document is the anchor for all the enhancements, but it is necessary to deliver these differently depending on whether it is available in electronic form or is itself the original. In the latter case optical character recognition or some form of signature analysis can be used. At worst, pages can be identified by additional marking. *Glyph codes* [Hecht 1994] are a form of high density, unobtrusive bar-code that look particularly promising for this purpose.

Finally, the user interface is being extended to exploit the new facilities. This involves work on free-hand gestures [Baudel & Beaudoin-Lafon 1993], although without the requirement to wear a special dataglove. The idea is to move away from explicit commands from a menu and towards a scheme in which the camera observes the activities of the user and infers the appropriate assistance to offer. Simply picking up a pencil or eraser or reference book should cause the program to react.

CONCLUSIONS

This paper has reviewed some recent work on human-computer interaction involving the use of image capture and projected video in the office environment. The original motivation for this work was to experiment with new user interfaces that required no special knowledge to operate. In the event, a much more important result emerged, defining a new approach to computing in the workplace. Computational enhancements can be added to everyday objects such as paper on the desk or writing on a whiteboard; computing can be brought out of the workstation and into the real world.

This has raised a number of technical challenges to do with image processing, registration of real and synthetic worlds and the design of user interfaces. Some solutions have been presented and work is continuing. In particular, the application to multi-media publishing presents interesting challenges.

However, there are also social problems associated with this approach. The cumbersome paraphernalia of virtual reality makes it very obvious that the user is leaving the real world and entering a synthetic one. Bringing the computer out of the workstation and into everyday objects blurs this distinction and can be unsettling for users who no longer feel in command, as previously inanimate objects acquire lives of their own.

There are also obvious privacy concerns. If you do not need to use special tools to communicate with the computer, it must necessarily be aware of what you are doing all the time so that it can take appropriate actions when required. This may well be information that users would not want to broadcast publicly. We are investigating these issues further as well.

ACKNOWLEDGEMENTS

The DigitalDesk was built by Pierre Wellner and BrightBoard by Quentin Stafford-Fraser, both research students sponsored by Rank Xerox in the Computer Laboratory at the University of Cambridge. Current work on *animated paper documents* is sponsored by the EPSRC under grant GR/J65969. Many colleagues in the Computer Laboratory and at RXRC Cambridge (EuroPARC) have contributed to the work described here: Kathy Carter, Steve Freeman, Mik Lamming, Wendy Mackay, William Newman and David Wheeler deserve special thanks.

REFERENCES

Thomas Baudel & Michel Beaudoin-Lafon: *Charade -remote control of objects using free-hand gestures*, Communications of the ACM 36(7), July 1993.

Kathy Carter: *Computer aided design - back to the drawing board*, Proceedings of the International Symposium on Creativity and Cognition, Loughborough, April 1993.

Steve Freeman: *An architecture for distributed user interfaces*, Technical Report 342 (PhD dissertation), University of Cambridge Computer Laboratory, July 1994

David Hecht: *Embedded data glyph technology for hardcopy digital documents*, Proceedings of the SPIE Color Hard Copy and Graphic Arts Conference, February 1994.

Wendy Mackay, Gilles Velay, Kathy Carter, Chaoying Ma & Daniele Pagani: *Augmenting reality - adding computational dimensions to paper*, Communications of the ACM 36(7), July 1993.

Greg Nelson (editor): *Systems programming with Modula-3*, Prentice-Hall, 1991.

William Newman & Pierre Wellner: *A desk that supports computer-based interaction with paper documents*, Proceedings of the ACM Conference on Human Factors in Computing Systems, Monterey, May 1992.

RXRC: *The EuroCODE project*, Rank Xerox Research Centre Cambridge, June 1993.

Quentin Stafford-Fraser: *BrightBoard - a video augmented environment*, Technical report EPC-1995-108, Rank Xerox Research Centre Cambridge Laboratory, March 1995.

Pierre Wellner: *The DigitalDesk calculator - tangible manipulation on a desk-top display,* Proceedings of the ACM Symposium on User Interface Software and Technology, Hilton Head, November 1991.

Pierre Wellner: *Tactile manipulation on the DigitalDesk,* Video presentation at the ACM Conference on Human Factors in Computing Systems, Monterey, May 1992.

Pierre Wellner: *Interacting with paper on the DigitalDesk*, Technical Report 330 (PhD dissertation), University of Cambridge Computer Laboratory, March 1994.