MixPad: Augmenting Interactive Paper with Mice & Keyboards for Cross-media and Fine-grained Interaction with Documents

Xin Yang[1], Chunyuan Liao[2], Qiong Liu[2]

[1]Dept. of Electrical and Computer Engineering, University of California, Santa Barbara, CA, 93106, USA

[2]FX Palo Alto Laboratory, 3174 Porter Drive, Palo Alto, California, 94304, USA

xinyang@umail.ucsb.edu, {liao, liu}@fxpal.com

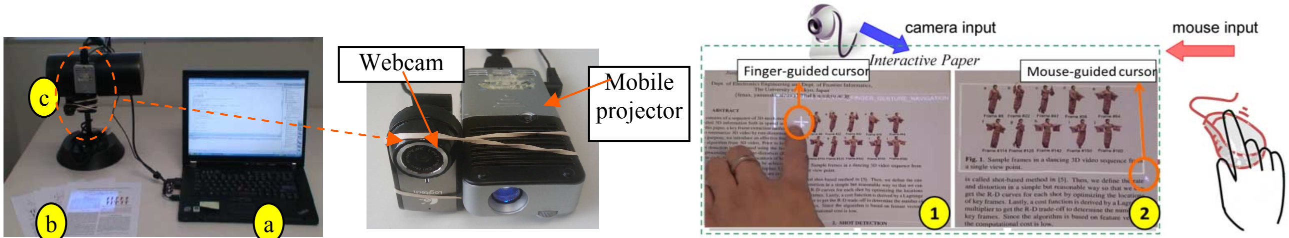

Figure 1. (left) MixPad interface components: (a) laptop, (b) printout and (c) camera-projector unit. (middle) Close-up of the camera-projector unit. (right) Steps to select a figure on paper: (1) point a finger to it, and (2) precisely select it using a mouse

ABSTRACT

Existing interactive paper systems suffer from the disparate input devices for paper and computers. The finger-pen-only input on paper causes frequent devices switching (e.g. pen vs. mouse) during cross-media interactions, and may have issues of occlusion and precision. We propose MixPad, a novel interactive paper system, which allows users to exploit mice and keyboards to digitally manipulate fine-grained document content on paper, such as copying an arbitrary image region to a computer and clicking on a word for web search. With the combined input channels, MixPad enables richer digital functions on paper and facilitates bimanual operations cross different media. A preliminary user study shows positive feedback on this interaction technique.

Categories and Subject Descriptors

H.5.2 **User Interfaces**: **Interaction Styles**

Keywords: Interactive Paper, Mouse, Keyboard, Cross-media, Fine-grained, Bimanual, Document

1. INTRODUCTION

Paper is still extensively used in conjunction with computers, due to its high display quality, tangibility, robustness and flexibility in spatial arrangement, which existing computers can hardly match. Nevertheless, paper lacks computational capability that computers offer. To combine the complementary advantages of paper and computers, many interactive paper systems [1,3,4,6] have been proposed to facilitate mixed use of paper and a computer side by side on a table (e.g. Figure 1). With this setting, users interact with paper documents with fingers or pens, and perform cross-media operations, e.g. copying a figure from paper to a laptop [3,4].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'12, October 29–November 2, 2012, Nara, Japan.

Copyright 2012 ACM 978-1-4503-1089-5/12/10...$15.00.

However, this paper-computer federation has to face a severe problem that the disparate input devices of the two different media cause inconvenient device switching during interactions cross paper and computers. Imaging a user writing PowerPoint slides on a laptop, to copy a graph from a printed reference article to the slides, the user may have to first drop her mouse, switch to a finger or a pen to select the targeted content, and switch back to mouse operation again. Moreover, she might also need to adjust body pose (e.g. towards paper vs. towards a screen) for the task. Such overhead of device switching and body adjustment breaks the user perception of a continuous document workspace spanning the paper and screen.

Further, although the conventional finger-based interaction works fine for normal interactive paper tasks, they are not geared toward manipulating *fine-grained* content on paper, such as individual words or specific image portions. The precision and performance of the finger input on paper [1,3,6] are limited, and even worse for distant targets out of the immediate reach of users' hands. Lastly, either finger or pen input does not support efficient, compact and large amount of text annotation on paper.

To address the above issues on cross-media and fine-grained interaction with paper document, we propose a novel interactive paper system called MixPad. Similar to many vision-projection based systems [1, 3, 4, 6], MixPad consists of a laptop, ordinary printouts and a camera-projector unit which recognizes the printouts and detects pen and finger gestures (Figure 1). Nevertheless, distinguished from those systems, MixPad adopts mice and keyboards as two additional first-class input channels.

As illustrated in Figure 1, to copy a figure from paper to PowerPoint slides, the user first roughly points her left index finger to the figure, without dropping the mouse from her right hand. In response, the system immediately moves and projects the mouse cursor around the finger on paper. The user continues to use her right hand to move the curse to select a rectangular region. Upon the selection done, the user presses CTRL+C short-cut keys with her left hand to copy the region, followed by mouse moving and pasting operations with her right hand. Moreover, the user can perform a continuous mouse drag & drop operation to link a web page on screen to a keyword on paper; she can also type text annotations that are then

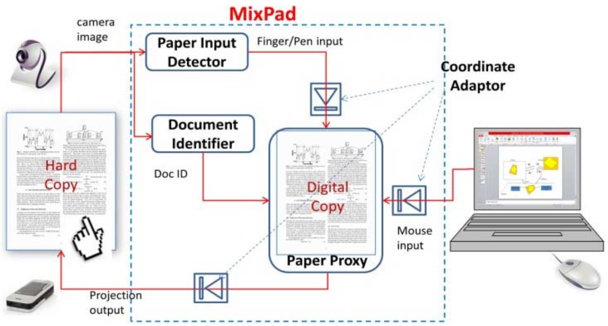

Figure 2. The architecture of MixPad system

attached to a specified anchor segment on paper. In this way, MixPad avoids device switching for cross-media manipulation, and achieves more precise and efficient mouse and keyboard operations on paper.

We were inspired by Hartmann's work [2], which augments an interactive table with mice and keyboards to manipulate pure digital content. MixPad instead augments paper documents for mixed-media interaction. MixPad also shares with Augmented Surfaces [5] the concept of continuous hybrid workspace, but advances it by supporting markerless tracking, bi-manual, mixed input (e.g. finger and mouse) and word/pixel-level operations on paper, which promotes a more computer-like user experience with physical objects and thus smoothes the workspace of hybrid computing devices and media.

We present below the system infrastructure, the detailed interaction technique design, and report an early user study, which suggests positive feedback from participants.

2. SYSTEM OVERVIEW

As illustrated in Figure 2, MixPad acts as a bridge between a physical and a digital workspace. It consists of four major components, namely *Document-Identifier*, *Paper-Proxy*, *Paper-Input-Detector* and *Coordinate-Adaptor*. Document Identifier exploits a natural feature-based document recognition algorithm (similar to [4]) to recognize and track paper documents via a web camera, avoiding any special markers on paper. Upon the recognition done, Paper-Proxy loads the corresponding digital copy of the recognized paper document, and executes the commands issued through the mixed input from fingers, pens, mice and keyboards. The finger and pen inputs on paper are detected by Paper-Input-Detector based on their distinguishing colors from background, and passed to Paper-Proxy. The mouse and keyboard inputs are redirected from the laptop. In return Paper-Proxy generates visual feedback via a mobile projector. Consequently, it behaves like a proxy of the physical documents and interacts with other computing devices such as the laptop.

Coordinate-Adaptor plays a key role in integrating the mixed-media

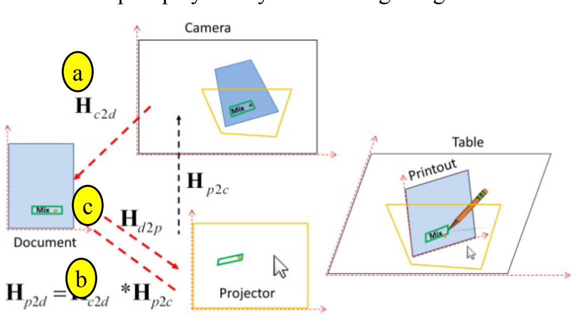

Figure 3. Transforms by Coordinate-Adaptor

input and output for paper interaction. It works between Paper-Proxy and other components, adapting all input/output to/from the uniformed Paper-Proxy coordinates (Figure 2). More specifically, for paper input, as illustrated in Figure 3-a, Coordinate-Adaptor translates the detected finger and pen tip positions from the camera image coordinates to the digital document coordinates, using the homographic transform $\mathbf{H}_{c2d}$ derived from the feature correspondence between the camera images and the digital document copy.

The mouse input is handled by the Coordinate-Adaptor if the cursor is within the display range of the projector, which is connected to the laptop as a secondary monitor via a VGA port. In this case, the cursor is projected onto paper documents, allowing users to manipulate paper content directly. For this purpose, Coordinate-Adaptor maps the cursor position from the projector coordinates to the digital document coordinates (Figure 3-b), via a homographic transform $\mathbf{H}_{p2d} = \mathbf{H}_{c2d} * \mathbf{H}_{p2c}$. $\mathbf{H}_{p2c}$ is the homographic transform from the projector coordinates to the camera coordinates, which is fixed through interaction sessions and previously calculated during system calibration.

The keyboard input is re-directed to Paper-Proxy if the projector display is on mouse focus. Once the paper-related commands are carried out within the digital copy, Coordinate-Adaptor then translates the resulting visual feedback from the document coordinates to the projector coordinates via another homographic transform $\mathbf{H}_{d2p} = \mathbf{H}_{p2d}^{-1}$, as shown in Figure 3-c.

By redirecting mouse input to paper, MixPad effectively addresses the issues of device switching in pen-based interactive paper systems such as [4], in that one can interact with paper document content without switching to a pen. Moreover, compared to finger - based interactive paper systems like [1,3,6], the MixPad mouse input enables operations at much finer granularity (i.e. pixel-level). Although there exist lots of fine-grained touch interaction techniques for screens (e.g. call-out on iPads), they do not work well for projector-based interfaces due to the occlusion of the projected call-out and the underlying document content. The redirected mouse input in MixPad has no such issue, does not demand extra sensing device like a capacitive touch pad, has higher input sampling rate than camera (100 Hz of mouse vs. 30fps of camera), and is very robust.

3. INTERACTION TECHNIQUES

With the supporting infrastructure, we focus on the paper + laptop setting as shown in Figure 1 and propose a set of interaction techniques to facilitate such mixed-media document manipulation.

3.1 Fine-grained Mouse Interaction on Paper

By mapping mouse pointing on a paper document to the equivalent operation within the digital copy of the document, MixPad enables computer-like mouse interaction on the physical surface. As demonstrated in Figure 1, a user can move the mouse to draw a marque selector to select a rectangular region on paper at pixel granularity. For best quality, MixPad does not directly use the selected region from the camera images, rather extracts the region from the corresponding high quality digital copy of the document based on the coordinate transform.

This idea can be extended to selection of text at word or symbol level. When the cursor moves over a word on paper, the text and boundary of the word are found from the digital document, and in return the projector highlights the word immediately for visual feedback to the user (Figure 4-3).

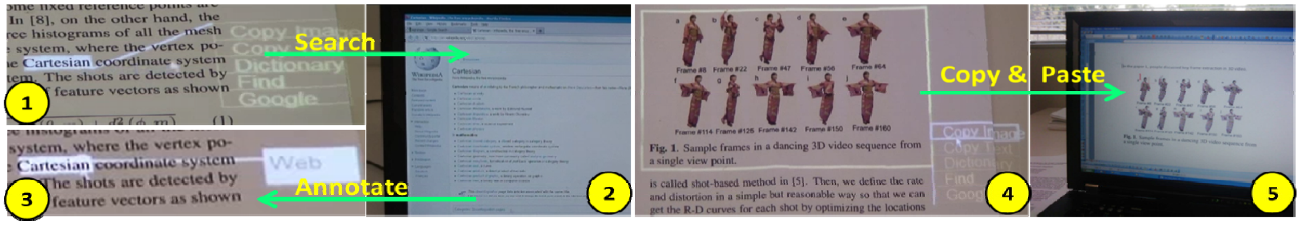

Figure 4. Cross-media interaction. (1) Google a word "Cartesian", (2) a found relevant Wikipedia web page, (3) the web page URL inserted as an annotation for the word "Cartesian", (4) (5)Copy and Paste from paper to the laptop

Context menus also work with mice on paper. MixPad detects the distribution of corner points in camera images, and locates a blank area around the selected object to project a context menu (Figure 4-1&4), from which the user can choose with mice a command to be applied to the selection. As a result, the user-familiar mouse operations can be readily migrated to paper documents, putting paper and computers on more equal footing.

3.2 Bimanual Mixed Input on Paper

MixPad does not discard the traditional finger interaction on paper, rather combines it with the mouse input for better user experience. Imaging a user switching the current working document from a PPT on the laptop to a printout, she might have to use the mouse to move the cursor all the way from the screen to the projector. This operation might be not smooth because of the long distance movement.

In response, MixPad adopts a two-stage bimanual finger-mouse input scheme. On the first stage, the user simply moves her non-dominant hand to roughly point a finger to a target object on paper. Once the pointing finger is detected, the cursor is immediately projected on paper and follows the finger (Figure 1-3). A paper-touch or pause-and-time-out event can be used to trigger the second stage, on which the input source is switched to the mouse in the user's dominant hand for precise selection.

This design is actually a variant of Guiard's Kinematic Chain theory [7] in that the non-dominant hand sets up the reference frame and the dominant one does a refinement. The difference lies in that MixPad uses two different media for the two hands. In the similar way, other bimanual interaction techniques such as Toolglass [8] could be applied to MixPad.

The mixed-media input on paper has advantages over finger-only interaction. Besides the interaction granularity, we also found in pilot tests that, for finger-only paper interaction, the users are used to holding a paper document with the non-dominant hand and doing finger gestures with the dominant hand, for which they have to drop the mouse first. One-hand interaction with paper does not work well, as printouts might move with the touching fingers, if not being fastened by the other hand.

3.3 Cross-media Interaction Using Mice

Enabling mouse operations on paper can effectively facilitate cross-media document interaction with a single input device. This results in an efficient and smooth experience of a continuous document space. For example, when reading an article on paper, one can first use a mouse to select an unknown word (e.g. "Cartesian" in Figure 4-1) and specify the "Google" command on paper, and then continuously takes the mouse to browse the web search results on the laptop (Figure 4-2). This paper to screen switching has less overhead than existing systems [3,4].

The cross-media interaction in the reverse direction is also smooth. After finding an important web page related to the unknown word, the user can use the mouse to continuously drag and drop the webpage URL from the laptop to paper to create an annotation (Figure 4-3). The annotation is rendered as an icon projected in the paper margin, and can be revisited later with a finger/mouse click. Similarly, when editing a PowerPoint document on a laptop, one can select, drag and drop a figure from paper to a laptop, without switching devices (Figure 4-4,5).

3.4 Augmenting Paper Using Keyboards

A keyboard can be used to add high fidelity text information to paper documents. For example, one can select a document segment on paper and then type detailed text annotation for it. Compared to hand-written annotation, this method is faster for long text, and easier to be processed by computers for search. And rendering of the annotations is very flexible: they can be presented in a much more compact form than handwriting to save space; and the displayed detail level can vary with specific viewing modes.

The commonly used short-cut keys are also allowed in conjunction with mouse operations on paper. For example, one can first take a mouse to select a paragraph on paper, press ctrl-c to copy it, move the mouse to switch the input focus to a WORD document on the laptop, and press ctrl-v to paste the selected content. In this way MixPad effectively narrows the gap between paper and computers.

4. INITIAL USER EVALUATION

To understand the capability and limitations of MixPad, we conducted an informal user study with six colleagues (3 females, 3 males who are not affiliated with this project).

Tasks

We examined the user experience at two aspects: performance of object selection on paper and device switching during cross-media interaction. For the first one, we considered objects at different granularity levels: fine-grained objects like individual words and coarse-grained large objects like figures. For the second one, we considered two typical device switching sequences: paper-screen-paper and screen-paper-screen.

Accordingly, the study consisted of two cross-media tasks. The first task investigated the word-level object selection and paper-screen-paper transition. Participants were asked to select designated words and issue a "Google" command on paper, browse the results on the laptop, and then drag and drop the best webpage link onto paper. The second task examined the coarse-level object selection and screen-paper-screen transition. Participants were asked to initially work on a PowerPoint document on the laptop, select and copy designated regions of a paper document, and then paste them to the slides.

We looked into two conditions, namely the proposed bimanual mixed input and the traditional pen-only interaction on paper. In the pen-only condition, a normal blue tip pen was used. The system detects the pen tip from the camera images, and interprets that into equivalent mouse pointing events within the recognized digital document. We eliminated finger-only interaction, because it does not support well fine-grained interaction with the system setting, and

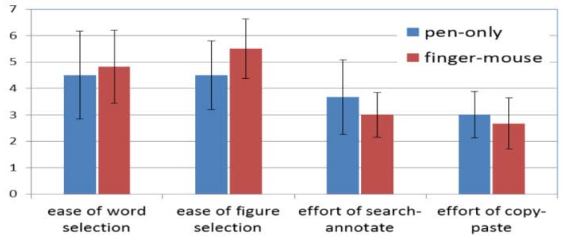

Figure 5. Averaged rating with 95% confidence intervals of finger-mouse input and pen-only input

neither two-handed nor one-handed finger interaction is suitable to the experiment tasks as explained in 3.2. Since the keyboard interaction is very simple, we skipped it in this test.

Consequently, we divided the whole test into four sessions; each session examines a combination of one interaction with one task. We counter-balanced the presentation order of the two interaction styles. There were no constraints on participants' body poses, hand and document positions.

Procedure

The study was conducted in a conference room. Each session began with a hands-on demonstration, followed by several practices and a six-trial testing. After a complete test, the participants answered a user experience questionnaire. Finally, we had an informal interview with the participants about their questionnaire responses. The test lasted about one hour. Since the test was an early stage evaluation, we did not quantitatively measure the user performance, but revolved around the subjective feedback and user behavior observation.

Results and Discussions

Overall, participants' reaction was positive and encouraging. They welcomed the idea of incorporating mice into paper and think it is helpful and convenient to use mice to interact with documents spanning paper and computers. Note due to the limited number of participants, we do not perform ANOVA on the responses and mainly reply on qualitative user feedback for just a rough trend, which we believe appropriate for such a preliminary test.

Pen-only vs. Finger-Mouse Interaction on Paper

On the whole, participants preferred using their fingers and mice to using pens for on-paper interactions. This result is attributed to four factors. First, relatively low frame processing speed and input sampling rate of pen input result in lag. Participants P3, P5 and P6 pointed out that lag of using pens annoyed them a lot during the content selection on paper. Second, the pen tip sometimes occludes the projected cursor, especially when interacting with distant objects. We observed that some participants needed to adjust their heads in order to see the cursor. To avoid such occlusion, some participants (P3 and P5) held the pen in the air, but this gesture is not natural. Third, pen input is not as accurate as mouse. P3 said "when using the mouse, the cursor can be moved per pixel; but pen could not do that". Fourth, timeout for pen operations is not intuitive and leads to inconvenience for adjusting the cursor. P6 said that she needed to place the pen correctly at the beginning, otherwise it always selected an unintended position if she pause a while for thinking. This is an implementation issue, and could be improved with some approaches such as using an infrared beam [1], but the fundamental issues of pen input like lags, occlusion, and lack of precision, are still hard to be solved.

The questionnaire responses also show such trend. When asked "how easier did you think you select a target word / figure?" the averaged participant responses on a 1~7 scale (7 means the easiest) was 4.83/5.5 for finger-mouse, compared to 4.5/4.5 for pen-only (Figure 5).

Pen-only vs. Finger-Mouse Interaction Cross Media

In general, the study suggests that participants preferred finger-mouse interaction for cross-media operations and thought it helps reduce physical efforts. For pen-only interaction, most participants (P1, P2, P3 and P4) only relied on one dominant hand to hold the input device, i.e. a pen or a mouse, which involves frequent device switch when performing cross-media tasks. Concurrently, their non-dominant hands were usually used to hold the paper to avoid undesirably movement. However, when using finger-mouse interaction, no device switch was needed in the dominant hand. Meanwhile, the non-dominant hand was free to be placed at any comfortable place after initializing a position on paper at start. This trend is also suggested by the questionnaire responses. When asked "how much coordination was required for cross-media search-annotation /copy-paste?" The averaged participant response on a 1~7 scale was 3/2.7 for finger-mouse compared to 3.7/3 for pen-only (Figure 5).

Furthermore, our original finger-mouse design requires participants to initialize a position on paper so that the system can project the cursor close to the target and help reduce the movement of mouse. However, some participants complained about the design in the scenario that the cursor is already projected on paper, which becomes unnecessary and misleading. To solve this problem, the future implementation will leverage smarter mode (finger vs. mouse) transition decision.

5. CONCLUSIONS AND FUTURE WORK

We present a novel interactive paper system called MixPad, which incorporates mouse and keyboard input into the finger-pen based paper interaction. This new interaction paradigm enables precise and high fidelity input on paper, avoids frequent input device switching for cross-media interaction, and therefore effectively bridges the paper and digital documents for a continuous document space. Future work includes real-time and robust finger detection to support "click" action and smarter finger vs. mouse mode transition for better user experience.

6. REFERENCES

1. Do-Lenh, S., F. Kaplan, A. Sharma, and P. Dillenbourg. Multi-finger interactions with papers on augmented tabletops. *Proceedings of TEI'09*, pp. 267-274.
2. Hartmann, B., M.R. Morris, H. Benko, and A.D. Wilson. Augmenting interactive tables with mice & keyboards. *Proceedings of UIST'09*, pp. 149-152.
3. Kane, S.K., D. Avrahami, J.O. Wobbrock, B. Harrison, A.D. Rea, M. Philipose, and A. LaMarca. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. *Proceedings of UIST'09*, pp. 129-138.
4. Liao, C., H. Tang, Q. Liu, P. Chiu, and F. Chen. FACT: fine-grained cross-media interaction with documents via a portable hybrid paper-laptop interface. *Proceedings of MM'10*, pp. 361-370.
5. Rekimoto, J. and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. *Proceedings of CHI'99*, pp. 378 - 385.
6. Wellner, P., Interacting with paper on the DigitalDesk. *Communications of the ACM*, 1993. 36(7): pp. 87 – 96.
7. Guiard, Y., Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. J. of MotorBehavior, 19 (4): 486-517, 1987.
8. Bier, E.A., et al. Toolglass and magic lenses: the see-through interface. *Proceedings of SIGGRAPH'93*, pp. 73 – 80.