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

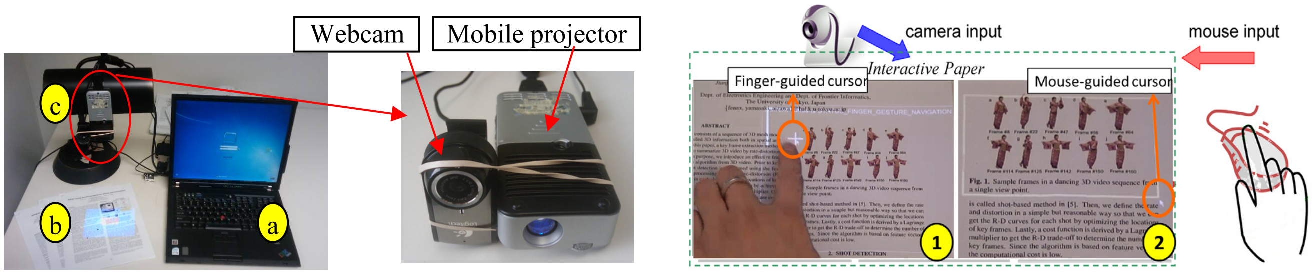

Figure 1. (left) MixPad interface components: (a) laptop, (b) printout and (c) camera-projector unit. (middle) Close-up of the camera-projector unit. (right) Steps to select a figure on paper: (1) point a finger to it, and (2) precisely select it using a mouse

ABSTRACT

Paper and Computers have complementary advantages and are used side by side in many scenarios. Interactive paper systems aim to combine the two media. However, most such systems only allow fingers and pens to interact with content on paper. This finger-pen-only input suffers from low precision, lag, instability and occlusion. Moreover, it incurs frequent device switch (e.g. pen vs. mouse) in users' hand during cross-media interactions, yielding inefficiency and interruptions of a document workspace continuum.

To address these limitations, we propose MixPad, a novel interactive paper system which incorporates mice and keyboards to enhance the conventional pen-finger-based paper interaction. Similar to many other systems, MixPad adopts a mobile camera-projector unit to recognize paper documents, detect pen and finger gestures and provide visual feedback. Unlike these systems, MixPad supports users to use mice and keyboards to select fine-grained content and create annotation on paper, and to facilitate bimanual operations for more efficient and smoother cross-media interaction. This novel interaction style combines the advantages of mice, keyboards, pens and fingers, enabling richer digital functions on paper.

Author Keywords

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

ACM Classification Keywords

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6): Interaction Styles

General Terms

Human factors, Design

INTRODUCTION

Paper and computers have complementary advantages and support each other in many scenarios. Paper is comfortable to read and annotate and flexible to arrange in space, but it lacks computational capability. Computers are powerful in computation such as document editing, archiving, sharing and search; nevertheless so far they can hardly match paper's high display quality, tangibility, robustness and flexible spatial arrangement.

To seamlessly combine the advantages of the two, many interactive paper systems [4, 5] focus on a mixed use of paper and a computer side by side on a table (e.g. Figure 1). Users can interact with content on paper using their fingers or pens and perform cross-media operations, e.g. copy and paste a figure from paper to a laptop [4].

However, most existing systems only use fingers or pens as input devices, which suffer from three problems. First, for on-paper interactions, the precision and performance of finger-pen-only input is limited, which is even worse for distant content out of the reach of user hands. Moreover, a real-time and robust finger/pen tip tracking is nontrivial. Some systems rely on extra sensing devices (e.g. a capacitive touch pad placed beneath a paper), yielding inconvenience. Other systems leverage camera-based tracking, however, they usually suffer from lag caused by frame processing and limited input sampling rate, i.e. 30fps. Second, for cross-media interactions which involve two different input devices, finger-pen-only input usually interferes with a convenient transition between the two media. For instance, when a user switches between paper and a computer she needs to change the input device (e.g. pen vs. mouse) in her dominant hand and adjust body pose (e.g. towards paper vs. towards a screen). Such device switch and body pose adjusting not only causes overhead, but also breaks the user perception of the continuum of a document workspace spanning paper and screens. Third, finger-pen-only-input does not support efficient, compact and large amount of annotations with detailed text on paper.

In this paper, we propose a novel interactive paper system called MixPad for more accurate, stable and faster on-paper interactions, and for more efficient and smoother cross-media transitions. Similar to many vision-projection based interactive paper systems [1, 3, 4, 6], MixPad consists of a laptop, ordinary paper documents and a camera-projector unit which recognizes the paper documents and detects pen and finger gestures (Figure 1). Distinguished from these systems, MixPad adopts mice and keyboards in conjunction with fingers. This is inspired by Hartmann's work [2], but differently MixPad augments interaction with paper while [2] focuses on interactive tables. In addition, MixPad facilitate bimanual interactions with information spanning paper and computers. As illustrated in Figure 1, to interact with a figure on paper, the user can first roughly point her left index finger to it. In response, the system projects the mouse cursor around the finger. The user then moves the mouse using her right hand to refine the selection.

For on-paper interactions, mice enable more precise input (i.e. pixel-level) than fingers and pens. In addition, mouse input does not demand extra sensing device (e.g. a capacitive touch pad) and has higher input sampling rate than camera (100 Hz of mouse vs. 30fps of camera). Moreover, mouse input is much more robust. For cross-media interactions, bimanual operation helps avoid input device switch in users' dominant hand and large body pose adjusting, thus resulting in a more efficient and smoother experience of a continuous document space.

An early user study on MixPad suggests that users welcome the idea of incorporating mice into interactive paper systems, and like manipulating documents bimanually with fingers and mice more than only using pens.

SYSTEM INFRASTRUCTURE

As illustrated in Figure 2, MixPad consists of four major components, namely *Camera Processor*, *Document Identificator*, *Paper-Digital Coordinator* and *Projector Processor*. The *Camera Processor* captures an image of a paper document, along with fingers. The captured document image is then identified by the *Document Identificator* as one of registered digital documents in the database. After that, the *Paper-Digital Coordinator* computes a coordinate transform between the captured image and its digital version. Based on the coordinate transform, associated digital information can also be precisely aligned with the paper document content. Meanwhile, on one hand, finger operations on paper are interpreted as equivalent mouse pointer manipulations on the digital version. On the other hand, mouse and keyboard input is interpreted as equivalent operations on paper. The resulting visual output for paper document, e.g. associated data and mouse cursor, is generated by the *Projector Processor* and projected on paper for direct visual feedback.

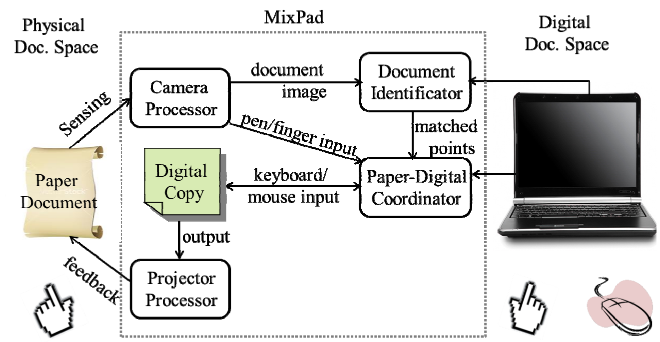

Figure 2. The architecture of MixPad system

INTERACTION TECHNIQUES

Progressive on-Paper Interaction using Mice

MixPad adopts a two-stage progressive finger-mouse operation for on-paper interaction. 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 projected on paper and follows the finger (Figure 1-(1)). In this way, the user can avoid using the mouse to move the cursor all the way from the screen to the paper. Since the rough pointing does not require high precision, it can be easily done by users and implemented with typical camera-based finger detection techniques.

A pause-and-time-out or paper-touch event can be used to trigger the second stage, on which the input source is automatically switched to the mouse, along with some auditory feedback. With the mouse, the user controls the cursor at high precision and performs user-familiar mouse operations on paper, such as a marque selection to select a figure (Figure 1-(2)). The idea can be extended to selection of any content on paper at different granularities, e.g. text at word-level or graphics and arbitrary regions at pixel level. This design is actually an implementation 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.

Context menus [4] 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 3-(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.

Smooth Cross-media Interaction Using Mice

Enabling mouse operations on paper can effectively facilitate cross-media document interaction, as users do not have to switch input devices in their dominant hand. This results in an efficient and smooth experience of a continuous document space. For example, when reading an article on paper, one can continuously use a mouse to select

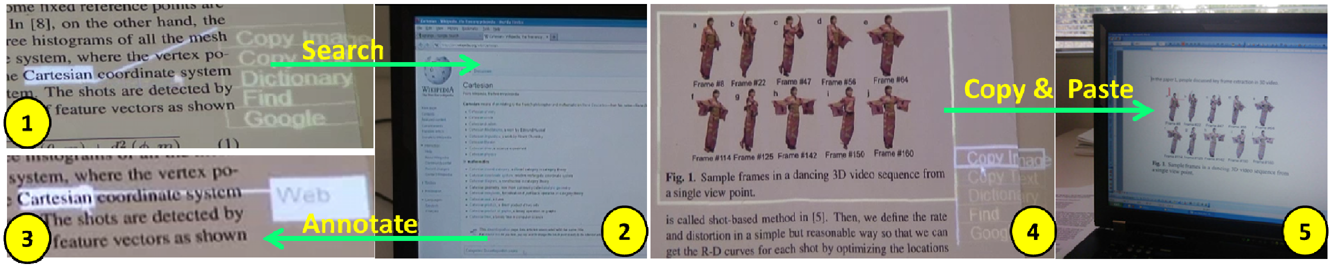

Figure 3. Cross-media interaction. (1) Google a word "Cartesian", (2) a found relevant Wikipedia web page, (3) the web page URL inserted as an annotation for the word "Cartesian", (4) (5)Copy and Paste from paper to the laptop

an unknown word on paper, "Google" it on a laptop, drag and drop a webpage URL from the laptop to paper to create an annotation. The annotation is rendered as an icon projected in the paper margin, and can be revisited later with a finger/mouse click. (Figure 3-(1) (2) (3)). 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 3-(4) (5)).

Augmenting Paper Interaction 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, our method is faster for long text and the result is easier to be indexed by computers and shared with other people. Moreover, the annotation can be rendered in a more compact form than handwriting to save space. Short-cut keys are also possible: one can press ctrl-c to copy a selected paragraph on paper, and ctrl-v to paste it into a WORD document on the laptop.

IMPLEMENTATION

We have fully implemented the MixPad system, written in C++/C#, running on a Windows XP laptop. A webcam of 640x480 pixels is connected to the laptop via a USB for capturing document and finger images; a mobile projector is also connected as a secondary display via a VGA port, for overlaying augmented information on paper. Currently, for fingertip detection, we adopt skin detection method based on a pre-trained skin model. For simplicity, we use 1 second timeout for input stage switching, mode switch and confirmation.

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 focus on examining the user experience at two aspects: performance of object selection for on-paper interaction and device switching during cross-media interaction. For the first one, we consider objects at different granularity levels: fine-grained small objects like individual words and coarse-grained large objects like figures. For the second one, we consider 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. In this task, participants were asked to select designated words and issue a "Google Search" commands on paper, browse the results on the laptop, and then drag and drop the best webpage link on paper (Figure 3-(1)(2)(3)). The second task examines the coarse-level object selection and screen-paper-screen transition. In particular, participants were asked to initially work on a PowerPoint document on the laptop, then select and "Copy" designated regions of a paper document and paste them to the slides (Figure 3-(4)(5)).

Furthermore, to compare the bimanual finger-mouse interaction and the traditional one-handed pen-only interaction on paper, all the participants were asked to complete the above two tasks using both interaction styles. We eliminate finger-only interaction due to its low precision arising from fat finger effects and instability with environmental changes that we found in pilot tests.

For each participant, we divide the whole test into four sessions; each session examines a combination of one interaction with one task. We also counter-balanced the presentation order of the two interaction styles. There were no constraints on participants' body poses, hand and document positions. Since the keyboard interaction is very simple, we skipped it in this test.

Procedure

The study was conducted in a conference room. Each session began with a hands-on demonstration, followed by several practices and 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, such as time cost for each interaction, but revolved around the subjective feedback and user behavior observation.

Results and Discussions

Overall, participants' reaction was positive and encouraging. They welcome the idea of incorporating mice and think it is helpful and convenient to use mice to interact with documents spanning paper and computers. Note as the limited number of participants, we do not perform ANOVA on the responses and mainly reply on qualitative user

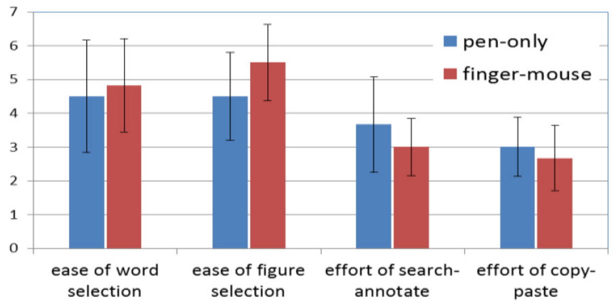

Figure 4. Averaged rating with 95% confidence intervals of finger-mouse input and pen-only input

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 prefer using their fingers and mice to only using pens for on-paper interactions. This result is attributed to four factors. First, time cost in frame processing and low input sample rate of pen-tip result in lag for on-paper interaction. 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 the pose of 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 very accurate, which makes it difficult for pixel-level selection. When selecting figures, 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 the time-out can be replaced 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 4).

Pen-only vs. Finger-Mouse Interaction across media

In general, the study suggests that participants prefer finger-mouse interaction for cross-media operations and thought it helps reducing physical efforts. This result can be explained by our observations. 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 move due to pen-paper friction all through the tasks. 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 4).

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.

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.

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): p. 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.