Interacting with Videos on Paper-like Displays

Roman Lissermann
Telecooperation Group
Technische Universität Darmstadt
lissermann@tk.informatik.tu-darmstadt.de

Simon Olberding
Telecooperation Group
Technische Universität Darmstadt
olberding@tk.informatik.tu-darmstadt.de

Max Mühlhäuser
Telecooperation Group
Technische Universität Darmstadt
max@tk.informatik.tu-darmstadt.de

Jürgen Steimle
Telecooperation Group
Technische Universität Darmstadt
steimle@tk.informatik.tu-darmstadt.de

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
Analog paper is still often preferred over electronic documents due to specific affordances and rich spatial interaction, in particular if multiple pages are laid out and handled simultaneously. We investigated how interaction with video can benefit from paper-like displays that support interaction with motion and sound. We present a system that includes novel interaction concepts for both video and audio. This includes spatial techniques for temporal navigation, arranging and grouping of videos, virtualizing and materializing contents, as well as focusing on multiple parallel audio sources.

Keywords
Flexible display, multiple displays, video, pile.

ACM Classification Keywords
H.5.2 [User Interfaces]:

General Terms
Design, Human Factors.

Introduction
Information on paper has one key advantage over today's computing devices: since paper is cheap, thin, and lightweight, we usually do not only use one single sheet, but we use multiple sheets simultaneously. Thereby we manipulate and organize information in

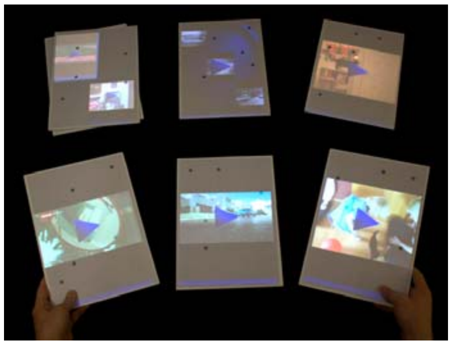


Figure 1: A system for interacting with multiple videos in physical space

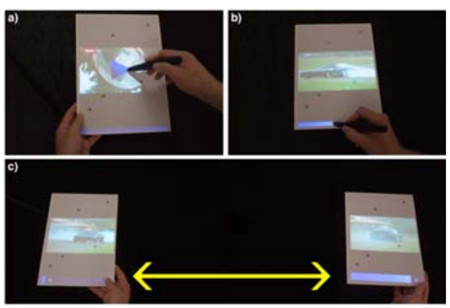


Figure 2: a) Playback, b) skimming, c) temporal navigation in space

physical space. This is useful for comparison, generating an overview and better orientation [4].

When contrasted to the wealth of well-established and effective physical interactions [8], the way we interact with videos today seems quite restricted. For more than hundred years of history, we are used to devices that display only one single movie at a time. In addition, while standard navigation techniques for videos have their obvious benefits, they lack the richness of physical interaction [4,5,8]. We argue that current and upcoming computing devices can do better and propose a different paradigm for videos. This consists of using many videos on many thin and lightweight displays simultaneously, similarly to how we lay out multiple printed documents on our desk.

With this concept, we aim at supporting three main scenarios of video usage, which we believe will benefit from multiple video displays:

1. Active video work: *Active reading* [1] is a well-studied domain. Active reading involves intensely engaging with documents, for instance by annotating, following references, and comparing documents. Often people work with multiple documents simultaneously and effectively arrange them in physical space to support their reading. Analogously, we propose active video work as a way of actively working with *video* material: people study a set of videos, prioritize content, find related contents, compare and (re)structure contents. We are convinced that also in the case of videos, the use of space provides effective support.

2. Playful video exploration: Multiple displays prove beneficial for exploring collections of videos, for instance at installations in places like shops, museums, or exhibition booths. We envision videos spread on booths or tables, where visitors can step by and playfully explore new topics or products individually or collaboratively. Hereby the system should be intuitively usable to support a playful exploration by people of all ages and professional backgrounds.

3. Video editing: Simple video editing is a common part of using videos as a consumer. For instance, people select and combine important sequences of a video in a personal excerpt. This simple editing is opposed to professional video editing, which is done during the production of videos. Previous work has studied how physical space can be used for video editing. It used physical objects like paper [5] and small active displays [9] as physical placeholders for a video. However the actual video is made available on a separate screen. In contrast, we propose using paper-like displays not only as tokens, but as the actual representation of the video, including audio output.

In this paper, we investigate how interactions known from physical documents can be transferred to the world of videos and fitted to a computing device that offers multiple thin electronic displays. Our work is inspired by PaperWindows [3], which addressed documents. We contribute a set of interaction techniques for time-continuous video content that take advantage of the characteristics of dynamic displays and go beyond established physical interactions such as arranging and piling of paper. The techniques allow users to play back and navigate through videos and collections of videos (see Fig. 1). They effectively support getting an overview of multiple videos as well as structuring and organizing video contents by

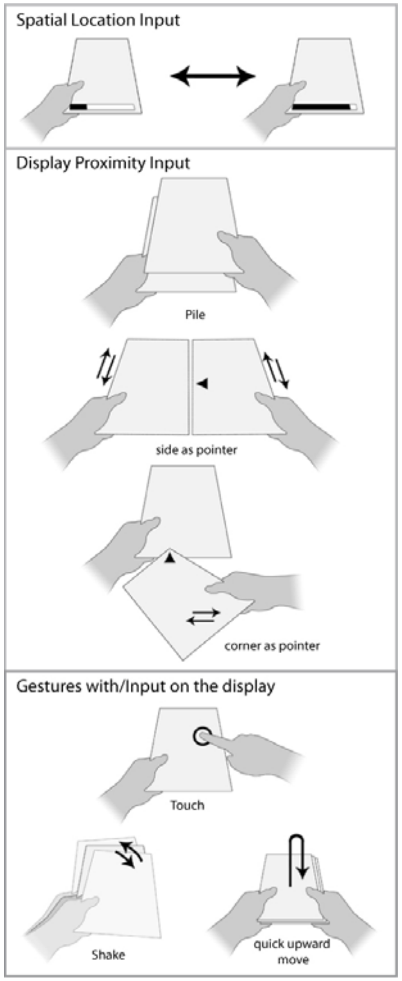


Figure 3: Input design space of multiple location aware displays

leveraging physical arrangements. Furthermore, we introduce interaction techniques that support the management of multiple displays by "virtualizing" content from displays and "materializing" virtual contents onto displays. To address the possibility of viewing multiple videos in parallel, we finally contribute several novel sound concepts that allow the user to focus on multiple audio sources of videos simultaneously.

Input design space of multiple displays

In this section, we systematically investigate the design space of input capabilities with location-aware multiple displays. We distinguish three different basic forms of input (see Fig. 3):

Spatial location Input: The absolute position of a display in physical space is captured. Input is provided by moving the display in space.

Display Proximity Input: The relative position of several neighboring displays in physical space is captured. Input is provided by piling displays or by using one display as a pointer to select content on another display.

Gestures with/Input on the display: The user can perform physical gestures with one display or with a set of displays, e.g. shaking. Moreover, the user can directly interact with contents on a display using direct touch or pen input (our prototype currently supports only pen input).

Interacting with Videos in Physical Space

In this section, we present interaction techniques that support a set of base activities for individual videos and collections of videos. These techniques leverage the manipulation and arrangement of one or several displays in physical space.

Temporal Navigation

Temporal navigation within a video is required for getting an overview of the video as well as for quickly accessing specific passages. Similarly to existing user interfaces for desktop computers and mobile devices, our design allows users to start, pause, and skim a video by directly interacting with widgets on the display using a stylus (see Fig. 2a, 2b). In contrast to most existing interfaces, it is possible to play back multiple videos simultaneously.

Space is a strong cue for encoding information, for instance sequences [4]. This motivated us to design a technique which in which the physical workspace encodes temporal positions. The timeline of the video is virtually spread out in physical space, extending from left to right within the user's arm reach (see Fig. 2c). Each spatial position is mapped to a temporal position within the video. By moving a display through space, the user navigates through the video. This technique allows for quickly skimming as well as for jumping back and forth between several passages of a video. To avoid interfering with free arrangements of multiple displays (see next subsection), temporal navigation is activated in a higher level above the table surface.

Arranging

Similar to arranging objects in the real world, the thin and lightweight displays can be freely arranged on the table surface. To state only some examples, two videos can be compared by placing them side by side. Multiple videos can be ordered in a spatially-encoded sequence. Videos can be prioritized by placing them closer or

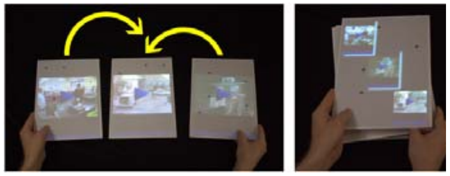


Figure 4: Physical piling of videos. The topmost display allows for interacting with all videos.

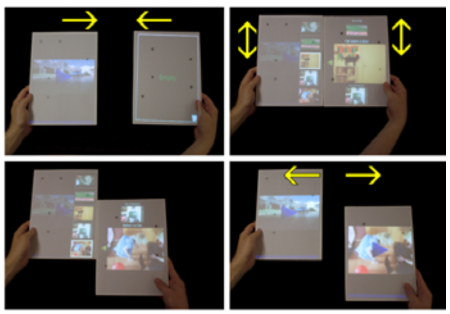


Figure 5: Accessing related videos by bringing displays side by side, selecting a related video and moving the displays apart

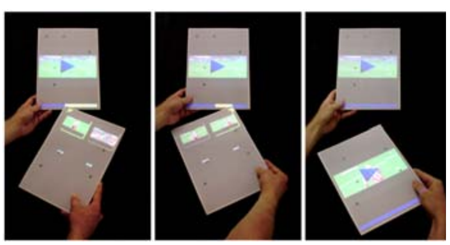


Figure 6: Playful video cutting with multiple displays

more distant to the user. Such arrangements enable powerful ways for organizing information in space [4].

See through pile

Previous research has shown the relevance of piling [6]. Users can place multiple displays on top of one another to form a pile of videos. Our pile is more advanced in comparison to piles of ordinary physical objects. Since the system is aware of which displays are occluded, the content of the entire pile is visualized on the topmost display, re-sulting in an "x-ray style" view (see Fig. 4). All this content on the topmost display is fully interactive. So the user can view, play or skim any video in the see through pile easily. Piling or unpiling does not interrupt playback; the video continues to play inside or outside a pile.

Accessing Related Videos

Many videos are organized in collections, in which they are linked to related videos. This is the case for such influential video platforms as YouTube or iTunes U. We present a spatial technique for navigating video relations using multiple displays. By bringing an empty display near to a display with a video (see Fig. 5), the mode for selecting related videos is entered. A list of related videos is visualized on the video. By moving the empty display or the video display up or down, a related video can be selected from the list. A preview of the currently selected video is then shown on the empty display. By moving one display apart, the related video is eventually selected and displayed on the previously empty.

We see mainly two advantages in this gesture while working with multiple displays. Firstly, the original video is not replaced by the related one, as in most current solutions, but remains visible. Hence a spatial overview can be easily generated by leaving the trace "where we came from". Secondly, multiple related videos can be opened by using multiple displays. These videos can then be spatially arranged and also viewed in parallel, if desired.

Playful Video Cutting

When people actively work with text documents, they highlight passages that are of high interest, write ex-cerpts, and create text collages by copying&pasting relevant passages into a new document. In contrast, video documents are usually consumed as is, without personalizing them. We propose a playful interaction technique for cutting videos. We do not aim for profes-sional video editing, but on providing a simple and playful interaction technique. This can be used for fo-cusing on specific passages of a video and for com-posing a "video excerpt".

For cutting out a section of an existing video, an additional empty display is needed. The corner of empty display is hereby used as a pointer device. By placing one corner of the empty display onto the timeline of the video, the corners can be used as a video cutting tool. The start and end positions of the cut are selected with the upper-left and upper-right corners, respectively. While selecting, the start and end frames are visualized on the previously empty display and the entire passage is highlighted in the video timeline (see Fig. 6). By moving the display apart, the cut is executed and the newly created video snippet is made available on this display. From now on, the user can interact with this snippet as with any ordinary video. The next section discusses how several physical video snippets can be combined to one video.

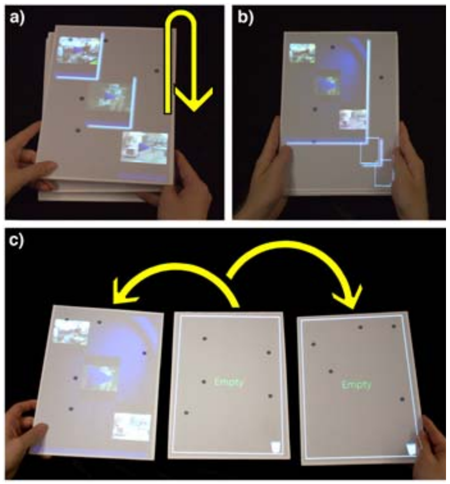
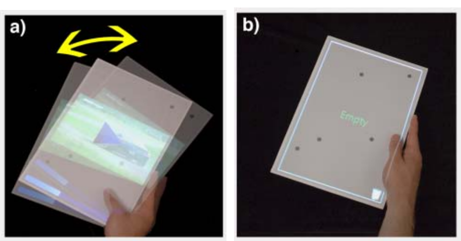


Figure 7: Virtualizing a physical pile (a), onto one single display (b). This clears the remaining displays (c).



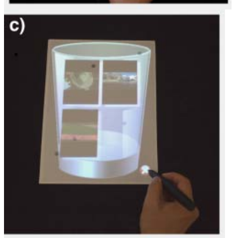


Figure 8: Clear content by shaking the display (a and b). Restore content from the recycle bin (c).

Managing multiple displays: Virtualizing and Materializing contents

Paper documents have a static mapping between contents and the physical carrier medium. One page of contents is permanently bound to one page of the carrier medium: this is a one-to-one mapping. Most computing devices have only one screen. Here, in contrast, contents are dynamic. A possibly infinite number of contents can be displayed on one physical screen: here we have a many-to-one mapping. Both types of mappings are well-understood.

Despite the expectation that future displays will be low-priced and lightweight, we imagine that users will have a number of displays that is much smaller than the number of sheets of paper that we use with printed documents today. Hence we cannot expect a tight one-to-one mapping of contents to displays. Also this would not be desirable, since it would run against the capabilities of displays to dynamically change their contents.

Therefore systems that offer many paper-like displays have a many-to-many mapping. Such systems afford physical interactions known from paper, however with a smaller number of carrier media. Previous work has shown how contents can be easily transferred from one display to another [3]. However, it is not clear how the handling and association of content on many displays should work. How can content be temporarily disassociated from displays to generate free carriers for displaying additional contents? How can such “virtualized” contents be “materialized” again and bound to physical carriers? We present interaction techniques that allow the user to combine contents onto one single display, distribute content over multiple displays, to clear and to restore content.

Combine And Distribute Content: By virtualizing videos that so far are bound to a physical display, physical displays can be freed from contents. Thereafter they can serve as physical carriers for additional videos. To virtualize one or several videos, the user creates a pile out of the respective displays. Quickly moving the entire pile upwards combines all videos in the topmost display. The remaining displays inside the pile become empty. The metaphor of this interaction is to push all videos up, which are caught by the topmost display.

The reverse direction, materializing videos from a video collection, is done by placing one or several empty displays underneath a display containing a video collection. By quickly moving the pile downwards, as many videos from the collection are materialized as empty displays are available (see Fig. 7).

Clear and Restore Content: Contents can also be virtualized by clearing a display. Clearing is performed by shaking the display, as if one shook contents off. Cleared contents are available in the recycle bin, which can be accessed by tapping on an icon that figures on all empty displays (see Fig. 8).

Sound concepts for Parallel videos

Since multiple videos can be laid out in physical space and played back simultaneously, multiple audio sources can be active in the same time. This produces the well-known cocktail party effect [2], which might make it difficult to perceive information conveyed on the audio channel. Browsing multiple audio sources by creating an audio-only environments have been investigated by Audio Hallway [7]. In our work, we refine this work for audiovisual contents and tangible interaction and introduce further spatial audio concepts.

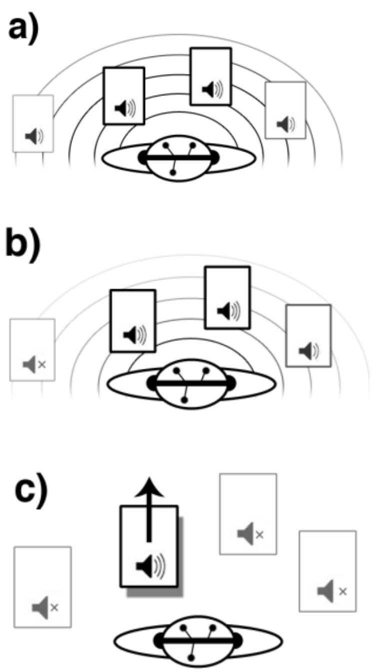


Figure 9: Sound concepts a) Real-World Behavior b) Distance-based Focusing c) Pick-up based Focusing

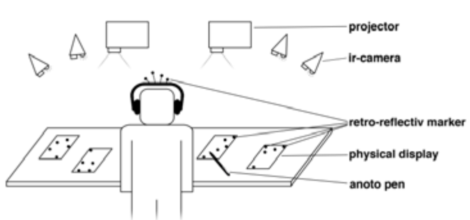


Figure 10: Technical setup

In this section we introduce two sound concepts that will allow users to more effectively focus on one or multiple sound sources that are located in space.

Distance-based Focusing: In work with paper documents it is well-established practice to focus on documents by placing them directly in front of the user. Documents are brought out of focus by placing them farther away, but still within an arm's reach [8]. Inspired by this behavior, we propose a sound concept for focusing on sound sources based on their distance.

Moving a display closer to or more distant from the user increases or decreases its volume (see Fig. 9b). In contrast to the Real-World Behavior, where distance has only a barely noticeable effect on the volume, the Euclidean distance between the user and the display is mapped inverse exponentially to the volume. As a consequence, volume can be finely adjusted. Placing the display an arm's length away is enough to reduce the volume to zero.

Pick-up-based Focusing: It is common practice to pick up an object for focusing on it. Inspired by this behavior, we introduce a further sound concept. Sound from displays that are lying down on the table is set to mute. By lifting up one or multiple displays, the sound of these displays is played back (see Fig. 9c). While picked up, sound sources expose a Real-World Behavior, being correctly located in space.

Implementation

Our prototype system realizes paper-like displays by tracking passive cardboards in real-time and projecting contents onto them. An overview of the system is shown in Fig. 10. For generating a 3D perception of sound, we used the OpenAL Framework. This setup can be used in fix installations, e.g. in libraries, shops or museums. Current advances in thin-film displays will soon render possible a mobile solution which consists of several wirelessly interconnected active displays.

Summary and Future Work

In this paper we have proposed a novel paradigm for spatial interaction with video. We have introduced a system that supports playback, flexible navigation and spatial organization of videos on multiple physical displays. We have presented interaction techniques that support the management of multiple displays. Furthermore, we have contributed a set of sound concepts that allow the user to focus on multiple audio sources of videos that play simultaneously. As future work, we plan to conduct a comprehensive user study.

REFERENCES

[1] Adler, M.J. and Van Doren, C.L. *How to read a book*. Simon and Schuster, 1972.

[2] Arons, B. A review of the cocktail party effect. *Journal of the American Voice IO Society 12*, 7 (1992).

[3] Holman, D., Vertegaal, R., Altosaar, M., Kl, C., Troje, N., and Johns, D. PaperWindows : Interaction Techniques for Digital Paper. *Interfaces*, (2005).

[4] Kirsh, D. The intelligent use of space. *Artificial Intelligence 73*, 1-2 (1995).

[5] Mackay, W.E. and Pagani, D.S. Video Mosaic: Laying Out Time in a Physical Space. In *Proc. MULTIMEDIA '94*.

[6] Mander, R., Salomon, G., and Wong, Y.Y. A "pile" metaphor for supporting casual organization of information. In *Proc. CHI '92*.

[7] Schmandt, C. Audio hallway: a virtual acoustic environment for browsing. In *Proc. UIST 98*.

[8] Sellen, A.J. and Harper, R.H.R. *The Myth of the Paperless Office*. The MIT Press, 2001.

[9] Zigelbaum, J., Horn, M.S., and Jacob, R.J.K. The Tangible Video Editor : Collaborative Video Editing with Active Tokens. *Design*, (2007).