EUROGRAPHICS 2007 / D. Cohen-Or and P. Slavík
(Guest Editors)

Volume 26 (2007), Number 3

Interactive Visual Workspaces with Dynamic Foveal Areas and Adaptive Composite Interfaces

Daniel Cotting and Markus Gross
ETH Zurich, Switzerland

Abstract

This paper presents novel techniques and metaphors for on-demand visual workspaces in everyday office environments, providing space-efficient, flexible and highly interactive graphical user interfaces using projected displays. For increased resolution, contents personalization and interactive visualization, the users can augment the large-scale projections with dynamic high-resolution foveal enhancements using a pocket light metaphor. To further optimize the presentation at a given resolution, the design of the displays can be modified interactively, and like a jigsaw puzzle, the layout can be customized using an adaptive compositing approach which supports free-form focus-and-context rendering. With a unified intensity-based tracking approach, we allow for natural multi-touch interaction with the information space through bare hands, pointers and pens on arbitrary surfaces.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces], I.3.3 [Picture/Image Generation], I.4.8 [Scene Analysis], I.3.6 [Methodology and Techniques], I.3.2 [Graphics Systems], I.3.1 [Hardware Architecture]

1. Introduction

Computer technology is increasingly migrating from traditional desktops to novel forms of ubiquitous displays on tabletops and walls of our environments. This process is mainly driven by the inherent limitations of classical computer and home entertainment screens, which are generally restricted in configuration and interaction possibilities.

Based on this observation, we envision a fully interactive, projector-equipped workspace which provides flexible global control of the projected light at any location of the environment. In an intuitive way and with a maximum degree of flexibility, on-demand visual displays can be instantiated on any surface within this space. Subsequently, by scanning the environment, the displays can dynamically adapt to objects and persons in a smart way.

In previous work, several of these aspects have been addressed by adaptive instant displays [CZGF05] and interactive display bubbles [CG06], which allow the users to define instant, freeform screens in an interactive way on arbitrary surfaces. Although the resulting display metaphor has been well accepted, it still contains various limitations. Most importantly, the display resolution is constrained by the setup of the projectors generating the large-scale workspace and the interaction is solely based on laser pointers.

This creates the need for more elaborate techniques in the context of interactive visual workspaces. In this paper we present a set of three novel contributions. First, we introduce a pocket light metaphor for achieving dynamic high-resolution foveal overlays that can be enhanced with varying personal information. Second, to further increase the legibility of the projected contents, we allow for enlarged displays by cutting unused parts and by compositing arbitrary portions of the contents in an adaptive and space-efficient manner. Third, the operations can all be performed using a novel unified tracking approach allowing for a natural interaction based on bare fingers, pens and pointers. We support both a single-handed and a two-handed interaction for a direct display manipulation.

The realization of our techniques builds upon previous work on display bubbles [CG06] and draws upon different technologies in a large variety of research areas, whose most relevant previous work is summarized in Section 2. A brief overview of the hardware system and its setup procedure is given in Section 3, followed by our pocket light metaphor in Section 4. Our adaptive composite user interfaces are presented in Section 5, and the interaction techniques in Section 6. We conclude the paper with a presentation of current results, a discussion and an outlook on possible future work.

2. Related work

Recently many realizations of interactive visual workspaces have been developed, inspired by Wellner's DigitalDesk system [Wel93]. While most systems provide a fixed and limited resolution, some attempts have been made to enhance the display quality in a foveal area. Baudisch et al. have introduced focus-plus-context screens, where a foveal LCD display is surrounded by a low-resolution context projection [BGS01]. Similarly, the Escritoire system [AR05] implements an interactive foveal display with two fixed projectors oriented towards a rectangular desk. Both approaches only provide a static foveal area whose size and position cannot be modified. To lift this limitation, Staadt et al. [SAKH06] allow a projected foveal area to be redirected using significant hardware resources like a computer-controlled, steerable mirror inspired by the everywhere display project [Pin01]. Due to the chosen setup, the distance of the projector to the surface cannot be modified, limiting the achievable resolution gain and restricting the users in their interaction possibilities. All of the aforementioned approaches do not scale well as they do not easily support multiple foveal areas. As an exception, DTLens [FS05] allows multiple users to zoom into certain parts of the display using rectangular stretching regions, thus improving the legibility in user-defined areas, however not improving the local physical resolution. As a common, privacy-related limitation, the presented approaches do not allow the users to project private foveal areas, whose content is not known to the main display system. Although augmented surfaces [RS99] provide private areas, the proposed hybrid workspace, which consists of traditional screens and projections, does not provide the precise calibration necessary for the seamless generation of foveal areas. By contrast, our novel approach based on a pocket light metaphor supports the scalable projection of private content in continuously calibrated foveal areas of flexible position and size, thus lifting the various limitations of the aforementioned approaches.

In order to create correctly calibrated projections, the positions and orientations of the mobile projectors need to be continuously determined. Previous work required a camera dedicated to each portable device and visual markers on an instrumented projection surface [RBvB*04, BBRF05]. To circumvent this intervention into the physical working space, approaches have been developed which replace the tags in the environment by a variety of embedded sensors or visible calibration markers on the portable devices themselves [CB06, BFC05]. The tracking approach we have devised does not require such special sensors or markers. Using a single fixed camera, it supports multiple unmodified, off-the-shelf display and projection devices.

Related to our adaptive composite displays, the WinCuts system [TMC04] and Stuerzlinger et al. [SCPR06] have proposed the removal of irrelevant parts of screens and the jigsaw-like recomposition of user interfaces. However, contrary to our approach, they do not support a cross-platform compositing and they do not provide focus-and-context techniques and an adaptive reshaping.

The interaction input in visual workspaces is generally achieved using cameras [KIN*05, RM97, MITS05], tracked devices [HBL*06, AR05, VLS02] or sensors integrated into the surfaces [DL01, Rek02]. Frustrated total internal reflection [Han05], as a very promising approach, allows for an intuitive multi-touch input using a special, instrumented table. Similarly, our technique, which is based on a unified intensity-based tracking approach, provides both single-handed and bimanual input for multiple simultaneous users with the advantage of not being restricted to rather small and special interaction surfaces.

3. System overview

Our interactive visual workspaces rely on projection technology, since currently no other technology provides a way nearly as competitive and effective to build flexible, large-scale environments. The scalable setup consists of several networked I/O units, which can be individually oriented to cover the desired working space, usually consisting of several desks or walls. The modules, which include a computer, a DLP projector and a camera, are synchronized using a common signal generated by a microcontroller. In our implementation, the tasks are split among the cameras as shown in Figure 1. An infrared grayscale camera is used for detecting the user interaction, and another grayscale camera for the tracking of the foveal areas of the pocket light metaphor. Besides the I/O units of the large-scale projection, additional wireless laptops with portable pocket projectors are used for creating the high-resolution foveal areas.

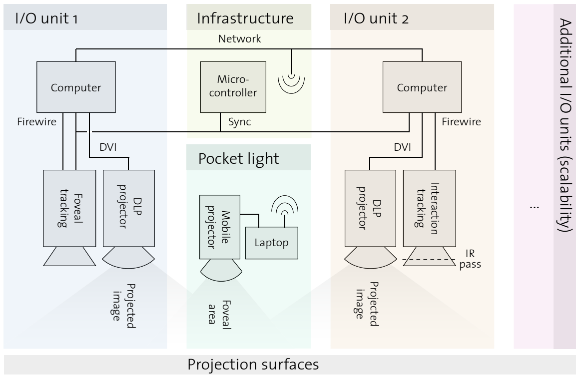

Figure 1: *Multi-unit configuration of the scalable setup used for realizing the interactive visual workspaces.*

To achieve a seamless alignment of our display projections, the devices of the I/O units must be calibrated intrinsically and extrinsically with relation to each other. To achieve this goal, we initiate a calibration for both the cameras and the projectors by an approach based on a propagation of the Euclidean structure using point correspondences embedded into binary patterns [CZGF05]. Using the same patterns, initially presented by Vuylsteke and Oosterlinck [VO90], we reconstruct the projection surfaces.

4. Pocket light metaphor

In order to locally increase the resolution of our projector-based workspace and to support the display of additional, personalized contents, we allow for dynamic foveal enhancements. For that purpose, we combine the low-resolution context information of the projection provided by the large-scale setup (cf. Section 3) with high-resolution focus areas overlaid by portable projectors. Recently, such projection units have continuously become smaller and lighter due to the constant miniaturization of consumer electronics. Thus, one can expect that tiny projectors will be integrated into a large variety of handheld devices like cell phones or PDAs in the near future, and therefore projection capabilities will quickly become ubiquitous. Similar to pocket lights, we can move the portable projectors, which represent nothing but more elaborate light sources, to dynamically move and resize the foveal areas. The surrounding large-scale projection is either blanked out in the corresponding areas, or it is maintained to achieve a blended imagery.

Note that contrary to existing approaches to foveal displays [SAKH06, BGS01, AR05], the focus area in our setting is not static and the projector's distance to the wall can be modified, giving the user an additional degree of freedom and also providing a varying resolution of the focus area. Therefore, potentially much more detail can be displayed if desired. Furthermore, the focus area can be personalized by the portable device without knowledge of the person's identity by the system, thus avoiding sensitive privacy issues. Additionally, the hardware complexity and cost is reduced compared to previous approaches supporting flexible foveal areas [SAKH06], since no computer-controlled motorized equipment is necessary.

4.1. Continuous calibration

To create a steady and well-defined overlay projection and a precise augmentation of the large-scale displays, a movable handheld projector must continuously be calibrated. For a reliable determination of the dynamic projection properties, we include a white border into the image stream of the portable device, as shown in Figure 2 a). This tracking frame can easily be captured by a precalibrated camera and allows the surface-to-device homography H_{SD} for piecewise planar surfaces to be computed on-the-fly during runtime. Whenever possible, we embed a frame of type A, depicted in Figure 2 b), which provides a larger projection area, but which does not provide information about the frame orientation, contrary to the frame of type B shown in Figure 2 c).

Border tracking. We determine the four corners of the projected tracking frames by following the recipe shown in Figure 3. Our pipeline consists of a sequence of well proven processing algorithms [Int06]. After performing an adaptive thresholding on the captured camera image, the connected components of the brightest areas are determined, and their polygonal contours are simplified using the Douglas-Peucker algorithm [DP73]. Subsequently, we accept all convex results containing exactly four points and having a predefined minimum area A_{min} in image-space, which reliably rejects wrong positives due to potential noise in the camera sensors. The four vertices of every accepted contour are then refined using a subpixel corner estimation. After reconstructing the corresponding points on the 3D surface, we try to track every contour over time by matching it to a previously detected frame using spatial 3D distance constraints. If a valid match is found, the previously determined orientation of the contour is propagated, else we try to determine

Figure 2: *Tracking border. a) Pocket light unit projecting a white tracking border. b) Sample image with surrounding tracking frame of type A. c) Sample image with surrounding tracking frame of type B, which also encodes the orientation.*

the orientation using a tracking frame of type B by comparing the brightnesses of the neighborhoods of the edge midpoints. If in such a case the currently displayed frame is not of type B, we notify the corresponding device to activate type B for future use and discard the corresponding contour. As a last step, all the remaining valid contours are Kalman-filtered [KB61] for increased resilience to jittering. Devices, whose contour has successfully been determined, can switch back to type A frames.

Superposition with large-scale projection. To avoid an interference of the projected tracking frame with the large-scale projection, our camera is synchronized with the fixed projectors [CWDG07] and acquires images during a blanked out interval of the modulation pattern of the DLP projectors [CNGF04], as shown in Figure 4.

Computation of homography. After processing all the pipeline stages described in Figure 3, the contour points are consistently sorted in a clockwise orientation starting from the top left corner of the projected content. Given the resulting

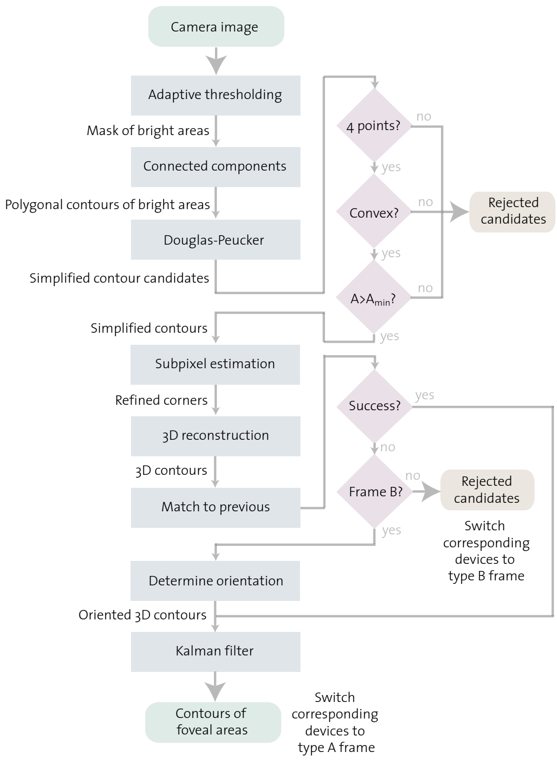

Figure 3: *Detection pipeline of the foveal tracking borders.*

homogeneous coordinates $P_{1...4}$ of the tracking corners in surface parametrization and given the horizontal and vertical projector resolutions Res_x and Res_y, we compute the surface-to-device homography H_{SD} such that

$$(0, 0, 1)^T \sim H_{SD} \cdot P_1,$$
$$(Res_x, 0, 1)^T \sim H_{SD} \cdot P_2,$$
$$(Res_x, Res_y, 1)^T \sim H_{SD} \cdot P_3,$$
$$(0, Res_y, 1)^T \sim H_{SD} \cdot P_4.$$

During projection, this homography can be used to warp the imagery defined in the surface parametrization to the perspective of the mobile projector, thus creating correctly calibrated foveal overlays.

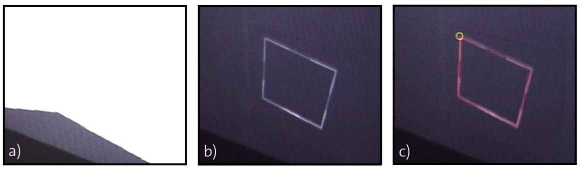

Figure 4: *Superposition of the tracking frame with the large-scale projection. a) The large-scale projection, which is set to display a uniform white surface in this example, interferes with the tracking border of the portable device. b) By exploiting the modulation pattern of DLP projectors, the large-scale projection can be rendered invisible to the camera. c) After processing the snapshot of the type B frame, the location and orientation of the foveal area can be determined.*

4.2. Interaction modes

Besides displaying fixed overlays in dynamic areas, we allow for varying content and an interactive data manipulation by additionally considering the change in distance of the portable projector from the surface, similar to the concept of information granularities by Cao et al. [CB06].

Activation of interaction. The user can intuitively switch between different interaction modes by tilting the projector in a predefined direction, as shown in Figure 5 a). Alternatively, the user can also rotate the inclined portable projector around a constant pointing direction, which is illustrated in Figure 5 b). The currently active mode is computed by considering the 3D shape of the projected tracking frame and by determining the direction of the major foreshortening effect. For that purpose, we compute the edge length ratio for each pair of opposite edges of the tracking frame. The mode with the largest ratio is activated if the value exceeds a certain threshold, else no mode is enabled and the projector reverts to a standard non-interactive projection.

Distance computation. The modification of the content in the various interaction modes is based on a distance estimation of the projector from the surface. As a simple and robust metric we compute $d = k \cdot \sqrt{A}$, where A is the area of the projected tracking frame in the 3D surface parametrization and k is a projector-dependent, user-defined scaling factor. In each interaction mode, the distance is visualized using a dynamically updated slider at the display border.

Cursor mode. If in cursor mode the distance d falls below a threshold d_{min}, the projector is operating in a cursor-controlling interaction layer and we assume that the user wants to use the projector as an input device. Therefore, we compute the center of mass of the projected frame as a cursor position and allow the user to move the application pointer similar to the work of Beardsley et al. [BBRF05]. Reducing the distance below $d_{click} < d_{min}$ generates a click to trigger menu items or to directly work on the contents of the visual workspace. Due to the visual distance notification, the user can at any time safely navigate through the contents without triggering unintended clicks.

Content mode. In content mode, the portable projector switches between different application-dependent data for multi-layer display. For instance, when visualizing volume data, the user can interactively slice through the data set by moving the projector. By interpreting a time-dependent 2D movie as a 3D volume, one can also slice through the video cube to render individual frames. With several overlapping displays, the content mode navigates through the stack of displays and expands individual layers.

Zoom mode. By modifying the distance of the projector the contents in the foveal area can be scaled interactively.

Blend mode. This mode changes the degree of blending between the large-scale projection and the overlay imagery.

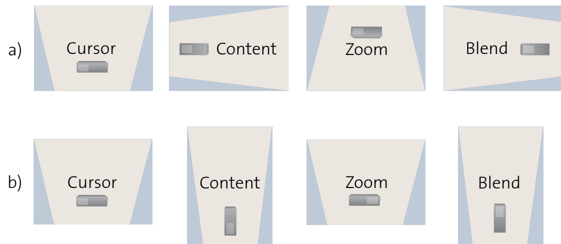

Figure 5: *Switching the portable device to an interactive mode. a) By tilting the projector downwards, to the left, upwards or to the right, the user can switch to different interaction modes. b) Rotation of the projector around a fixed pointing direction can alternatively be used to activate the various interaction modes.*

4.3. Hybrid displays

Since our tracking approach and the computation of the surface-to-device homography H_{SD} have been designed without any assumption about the display devices themselves, we can easily support alternative displays beyond portable projectors. A wide range of networked devices with built-in displays like notebooks, tablet PCs, PDAs and cell phones can be included into our framework to generate hybrid high-resolution foveal areas. Instead of applying the approaches presented in Section 4.2, the users can intuitively interact with various additional techniques (see Section 6).

5. Adaptive composite user interfaces

Besides proposing dynamic foveal enhancements based on a pocket light metaphor, we introduce composite user interfaces as an alternative solution to increase the legibility of the projected contents in a large-scale setup. The users can arbitrarily cut and zoom their displays, and subsequently the resulting pieces can be rearranged into novel, user-defined layouts. Contrary to previous approaches in cut user interfaces [SCPR06, TMC04], we allow for freeform cuts with surrounding context areas and we do not only offer cross-application compositions, but also cross-platform patchworks thanks to the protocol indirection layer which is used for the content generation [CG06]. Compared to the cascading, tiling or overlapping of traditional rectangular windows, our adaptive composite interfaces visualize more of the relevant information in parallel while additionally keeping helpful context data. Therefore, we can use the limited surface of the working space more effectively.

5.1. Focus-and-context warping

To compute our composite user interfaces, we extend a focus-and-context warping approach, which has been introduced by the display bubbles metaphor [CG06]. The original warping computation proceeds as follows: Given an arbitrary closed shape S, which defines the display portion of interest, a mapping $M\colon (x, y) \rightarrow (u, v)$ of the original rectangular screen content R is computed with the following two constraints: a) The defined focus shape S displays the enclosed content with a maximum fidelity. b) The remaining content $R \setminus S$ is smoothly arranged around the shape S in a context area C. The warping method constrains the mapping M to follow the field lines of a charge-free potential field defined on the projection surface by two electrostatic conductors set to fixed, but different potentials V_S and V_R, where one of the conductors encompasses the area enclosed by S and the other corresponds to the border of R.

5.2. Display composition

The original warping approach has been designed to individually warp single displays to freeform shapes with surrounding context areas. Since it treats each display separately, we need to extend the computation of the potential field to support multiple interacting displays as they occur in our user interface compositions.

Areas of interest. On every rectangular display R_i shown in the visual workspace, the user can define a focus area S_i by drawing a curve encompassing his area of interest, illustrated in Figure 6 a). Therefore, each display consists of a content relevant in the current work environment and a remaining part $R_i \setminus S_i$, which needs to be warped using our novel approach.

Collision avoidance. Prior to computing the warping, we ensure that no two areas of interest S_i and S_j, $j \neq i$ overlap in the 3D surface parametrization. We resolve the collisions as shown in Figure 6 b) by applying repulsion forces to the overlapping areas until no more collisions occur.

Adaptive reshaping. The space between the focus areas, which are now at their final locations, is used to display the remaining content $R_i \setminus S_i$ of each rectangular display R_i. The original warping approach considered the space that can maximally be used to correspond to the original rectangle R_i, hence the fixed potential V_R of a conductor positioned at the border of R_i. We modify the approach by additionally setting the potential field to a conductor of potential V_R at the locations, which are nearer to any other focus area S_j, $j \neq i$, as explained in Figure 6 c). Through this procedure, every display can maximally use the space which is nearest to its focus area. Thus, collisions of the context areas of the different components of the user interface compositions are avoided. In order to be able to trace the field lines anywhere within $R_i \setminus S_i$, as in the original approach, we create a small gradient at the locations that we have set to V_R by selectively computing a relaxation in the corresponding areas while keeping the rest of the potential field fixed. As a consequence, the final shape of the resulting display is pushed across all the areas initially set to V_R, therefore avoiding overlapping displays. For more details on the warping computation of the original approach, please refer to the display bubbles metaphor [CG06]. An illustration of two warped display shapes resulting from our modified procedure is depicted in Figure 6 d). As can be seen, our technique can be used to tile and arrange displays in a space-saving and resolution-enhancing manner while taking into account a user-defined information focus.

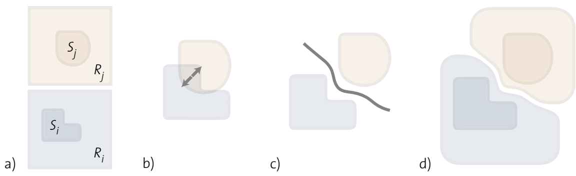

Figure 6: *Assembling adaptive composite interfaces. a) The user specifies the areas of interest on the rectangular displays and positions them in the visual workspace. b) If areas of interest collide, they are first separated using repulsion forces until no more collisions occur between the focus areas. c) While computing the potential field, we consider the distance of each point to the various focus areas, thus separating the entire space into subspaces associated to each display. d) Final resulting shapes of two assembled displays.*

6. Interaction

To accommodate multiple simultaneous users and to allow for an encumbrance-free and intuitive interaction, our visual workspaces can no longer rely solely on traditional input devices like keyboards and mice. For this reason, we have developed a set of flexible interaction techniques, including a projector-guided interaction (see Section 4.2), a tracking on surfaces with a unified intensity-based framework (Section 6.1), and a tracking in mid-air using morphological filters (Section 6.2). Our interaction scheme relies on a hand or pen tracking for proximity interaction, and laser pointers for distant interaction. In contrast to sensor-based surfaces [DL01, Rek02] or the tracking of digitizer or ultrasonic pens [AR05], we do not require any invasive or expensive equipment and allow for a more natural and intuitive interaction.

6.1. Unified intensity-based tracking

By detecting and tracking bright spots in camera images using an intensity-based framework, we can simultaneously support an interaction with laser pointers, light pens and bare hands in a unified approach.

Operating spectrum. To capture the user interaction only, and neither the large-scale projection nor the dynamic foveal overlays, we equip our interaction cameras with infrared bandpass filters with an approximate lower cut-off wavelength of 760 nm. As a consequence, the acquisition and the projection, which hardly emits any infrared light, practically operate in two distinct frequency bands (see Figure 7).

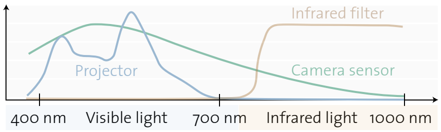

Figure 7: *The spectral distribution of the light emitted by a typical projector hardly contains any infrared light, while our camera can capture infrared, as shown by the spectral response of the camera sensor. By enhancing the camera with a filter whose transmittance is nearly zero in the visible spectrum, the camera can concentrate on infrared content while ignoring the visible projection.*

Fingers of bare hands. We cover our projection areas with a thin layer of infrared light using infrared laser plane modules. Whenever a person touches such a surface, light is reflected to the camera, therefore allowing an interaction with bare hands to be detected by our system (see Figure 8). Our modules are small, inexpensive and can easily be placed at the border of any ordinary surface which should become interactive. For immediate optical feedback to the user, we combine the imperceptible infrared laser module with a visible red laser diode.

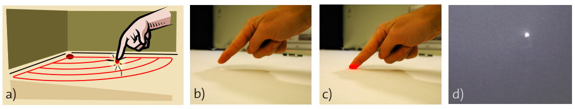

Figure 8: *Bare-hand interaction. a) A laser plane module covers the projection surface with a thin layer of infrared light. Whenever a finger penetrates the laser layer, light is reflected to the camera. b) Pointing slightly above the surface does not intersect the laser plane. c) Touching the surface intersects the light plane. d) The reflected beam is captured by the camera as a clearly visible bright dot.*

Light-emitting pens. As an alternative to a finger-based interaction, we support pens activating a light beam at their tip when touching the surface. Since most LEDs emit light in the infrared spectrum, a wide range of different pens are tracked in our unified interaction approach (see Figure 9).

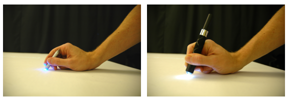

Figure 9: *Interaction with light-emitting pens on ordinary surfaces. The beam usually contains sufficient infrared light.*

Laser pointers. For interaction over distance, we propose laser pointers in both the visible and the invisible infrared spectrum. Even though common red laser pointers have a wavelength in the visible range around 650 nm, they are tracked by our infrared camera due to their very high intensity and the remaining transmittance of the infrared filters.

6.2. Mid-air tracking with morphological operators

In addition to supporting an interaction on the displays using our intensity-based framework, we detect bare fingers in mid-air, allowing the users to freely interact with nearby displays without touching any surface. Once again, by using infrared images, the projections are elegantly ignored by the cameras since the emitted light in the infrared spectrum is negligible. Inspired by Hung et al. [HYC*98], we detect fingers based on differencing and using morphological filters.

In our current implementation, the filters are applied as follows: Assuming that a finger has an expected maximum thickness of $2n$ pixels in the camera image, the background-subtracted snapshot is first dilated n times to smooth noisy segmentation contours, then eroded $2n$ times, and finally dilated n times again. We then determine and mark all the original foreground pixels, which have disappeared during this procedure (see Figure 10).

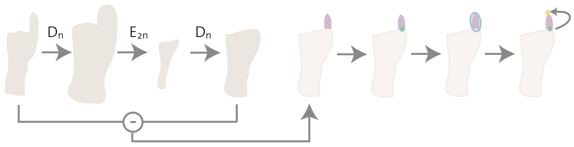

Figure 10: *Determining the foreground pixels disappearing after a sequence of morphological filters (purple). Further investigations yield the final pointer position (yellow).*

For each such pixel (x, y), we determine whether it is a neighbor of a pixel which has survived the morphological filters, i.e. we decide whether the pixel might belong to the base of a finger. If this is the case, we determine the connected component of the disappeared pixels which could potentially represent an eroded finger. To prevent further investigations of the same finger during the subsequent processing, we unmark all the corresponding pixels. In the next step we check whether the component considered has a size between $(4 \cdot n) \cdot n$ and $(4 \cdot 2n) \cdot 2n$ pixels. By defining these thresholds, we constrain the finger thickness to the range $[n, 2n]$, and the aspect ratio to 4:1 at the maximum. Subsequently, we perform a rough analysis of the shape by fitting an ellipse to the component (shown in blue in Figure 10), and if the aspect ratio of the resulting main axes is at least 2:1, we mirror the original pixel location (x, y) (shown in green in Figure 10) at the center of the connected component to get the final pointer position (shown in yellow in Figure 10). In our setting, this fast and efficient procedure has resulted in a very stable tracking of bare fingers.

6.3. Path interpretation

The interaction paths which are resulting from our two tracking approaches are all projected from 2D camera space to the 3D surface and then Kalman-filtered prior to being interpreted by our system. The computed positions can be used to generate a simple cursor input, or they can trigger both single-handed and two-handed gestures.

Individual paths. Individual interaction paths which are not part of a multi-touch input sequence are mapped to the appropriate displays, allowing the users to manipulate the screens and their content, or to trigger pie menus by activating and maintaining interaction at a stable location. To generate click events, the users quickly switch the interaction on and off in a sequence [CG06].

Coupled paths. When two interaction paths start almost simultaneously on an certain display, we assume that the user is manipulating the display with a multi-touch two-finger input (or with two pointers in parallel). We differentiate two cases: If the starting distance between the two paths is below a certain threshold, the user is performing a single handed scrolling operation as shown in Figure 11 a), else the user can stretch, rotate and move the entire display using two anchor points following his fingers as illustrated in Figure 11 b). Note that anytime multiple users can simultaneously interact on different displays in parallel, with both single-handed or two-handed interactions. In the case of multiple users, colored tokens in the visual workspace differentiate between the various concurrent cursor positions.

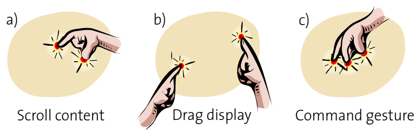

Figure 11: *Supported variations of multi-touch input.*

Hand gestures. When three or more fingers touch the display almost simultaneously in a spatial neighborhood as shown in Figure 11 c), we assume that the user is performing a single-handed command gesture. Sequences of tracked pointer locations are then transformed into the display coordinate frame and mapped to events using a previously trained neural network. The list of supported Graffiti-inspired gestures can be found in Figure 12. Whenever possible, we assign the gestures to the operations in an intuitive way, so for example a 'C'-shaped gesture activates a clone operation. Alternatively, the events can also be triggered using the pie menu or voice commands.

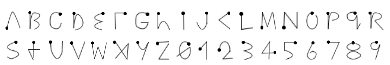

Figure 12: *Supported Graffiti-inspired gesture alphabet.*

7. Results

Based on the techniques and the metaphors presented in this paper, we have implemented an interactive environment providing flexible visual workspaces in a standard meeting room. Our setup operates with two ceiling-mounted I/O units (see Section 3), each consisting of a standard 2.8 GHz workstation with an NVIDIA Quadro FX3000G graphics board, a projector and a camera attached to an aluminum rig. We currently use Infocus X3 projectors with a resolution of 1024×768 pixels and Point Grey Dragonfly cameras delivering grayscale image streams with a 640×480 resolution at 30 frames per second. The system is complemented by a custom-made synchronization device (Toshiba TMP92FD54), a 100 Mbit/s ethernet hub (Netgear DS524) and a 54 Mbit/s WLAN access point for the portable display unit, which consists of a 1.7 GHz IBM Thinkpad T42 laptop and a Toshiba TDP-FF1A projector. For the unified intensity-based interaction, we are using standard off-the-shelf laser pointers, customized pens with an LED at their tip, and a laser plane generated by a 1 mW Picotronic LFD780-1-3 infrared laser, which is coupled to a KWB 0628-00 line-laser module of a visible red color for optical user feedback.

Note that the results related to the proposed interaction techniques are not presented separately, but are embedded into the following two sections dedicated to the pocket light metaphor and the adaptive composite interfaces. For a better impression of the interactive capabilities, please refer to the accompanying paper video.

7.1. Pocket light metaphor

To demonstrate the potential of our novel pocket light metaphor, we have implemented a collection of exemplary applications relying on mobile foveal areas. Flexible overlay projections allow for an interactive augmentation of geographic data (cf. Figure 13), and enable the enhancement of the content with user-defined information in a specific language (Figure 14), on information panels, schedules and maps. In these settings each user could carry his own portable projection device, e.g. embedded in a cell phone. The calibrated overlays are also suited for interactively visualizing medical data like x-ray imagery, and for overlaying and applying sketched notes (Figure 15). Augmentation of advertisements in public spaces or artistic applications can be envisioned as well. Most importantly, on-demand, the resolution of the imagery can be enhanced significantly (Figure 16). An example of a hybrid foveal setting using a personal laptop as an alternative display device is illustrated in Figure 17. Our direct, projector-based interaction techniques are presented in more detail in Figures 18, 19 and 20.

Due to the precise calibration of the system and the resilience to noise of the projector tracking, the overlays can be applied with a typical error of less than 4 mm on a 3.2 m × 1.6 m table surface, even when using a single camera covering an area of approximately 10 m^2 with a low-end 640×480 resolution only. In our setting, the processing pipeline of Section 4.1 requires around 8 ms for its computations, and therefore, the dynamic foveal areas can be tracked fully interactively. If desired, the perceived latency for smooth motions can be kept arbitrarily low by including the tracking predictions of the Kalman filter.

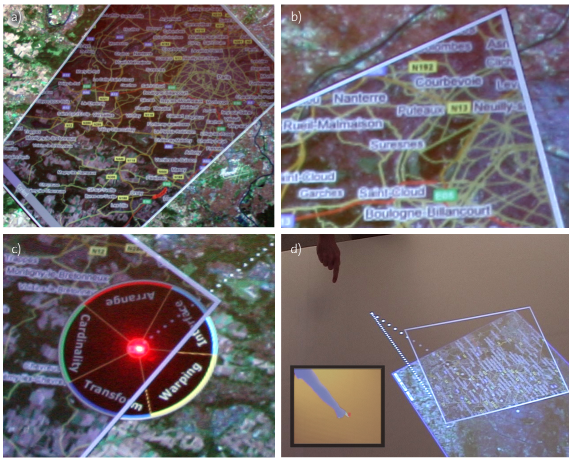

Figure 13: *Pocket light metaphor enhancing an aerial view with the corresponding street network. a) The foveal area can dynamically be moved on the large-scale projection. b) Close-up view. c) An on-screen menu triggered with a laser pointer affects both the large-scale display as well as the synchronized overlay projection. d) The user is moving the cursor with mid-air interaction to redefine the displayed content. The inset shows an extract of the segmented camera image highlighting the detected finger tip position.*

Figure 14: *Pocket light metaphor in multi-lingual environments. a) Overlaying English instructions on a Chinese directions panel. b) Displaying translated station names on a subway network overview. c) Replacing labels on a map.*

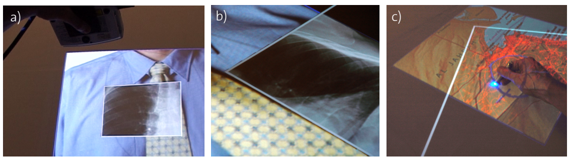

Figure 15: *a) By scanning an image using the portable projector, medical data like x-ray images can be visualized. b) Detailed view of the foveal area border. c) Annotations on a sketched overlay can be applied using an LED pen.*

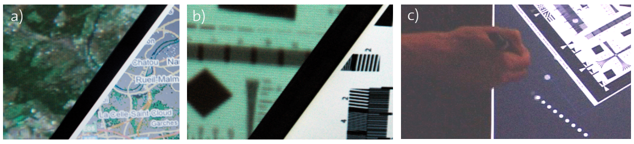

Figure 17: *Hybrid foveal displays. a) Customized foveal area using a laptop screen. b) Resolution comparison between the LCD panel and the surrounding large-scale projection. c) The user can directly work on the ordinary screen using a light-emitting pen.*

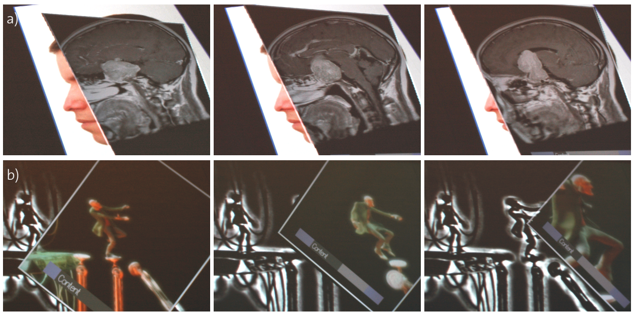

Figure 18: *Depending on its distance from the surface, the projector can switch between different data sets for multi-layer display. a) Different images of MRI volume data can be visualized. b) Video cubes can be sliced to render individual movie frames.*

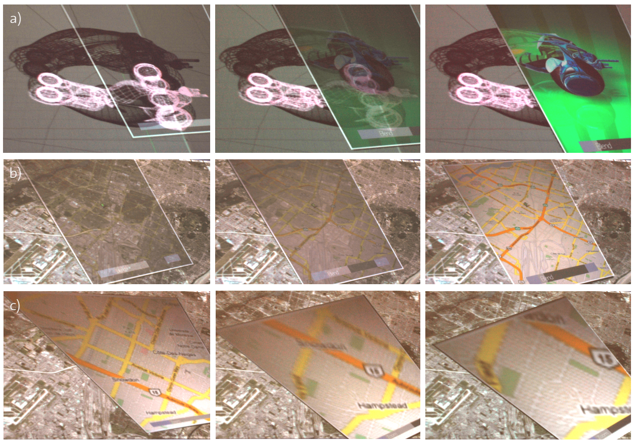

Figure 19: *Through projector interaction we allow for variable blending and zooming. a) A wireframe model is interpolated with its 3D rendering. b) A satellite image and its associated map are blended. c) The contents of the foveal area is scaled dynamically.*

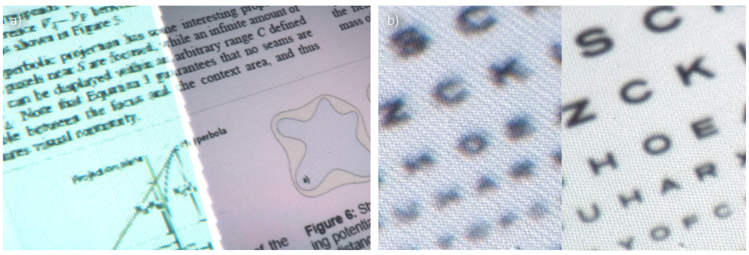

Figure 16: *The large-scale projection is enhanced by the potentially much higher resolution of our flexible foveal projections. a) Close-up view of the border between the foveal area and the surrounding projection. b) Side-by-side comparison between a large-scale projection and the foveal equivalent.*

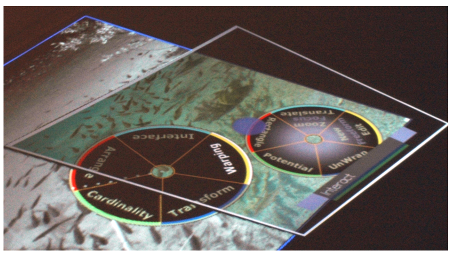

Figure 20: *The projector itself can serve as an input device when being moved within the interaction layer. A menu item can be selected and subsequently be applied to the display.*

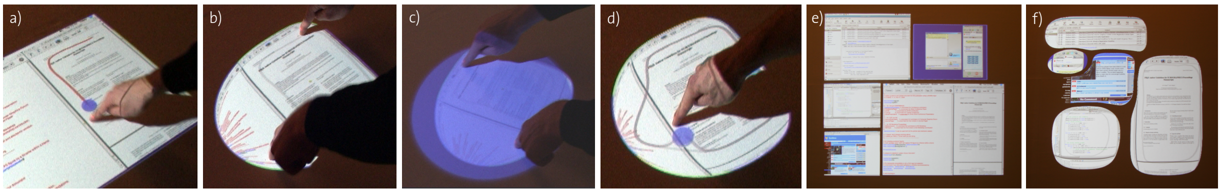

Figure 21: *Assembling a user-defined composite interface. a) Using bare fingers, the user defines areas of interest. b) With two-handed gestures, he can grab the displays. c) The individual pieces are dragged to their final locations. d) Anytime, the focus areas can be redefined. e) Rectangular starting layout. f) Resulting space-efficient composite interface with improved legibility.*

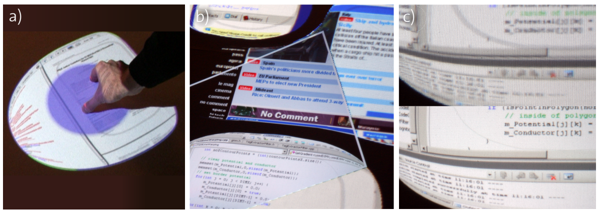

Figure 22: *The context areas of the adaptive displays can be explored with single-handed scrolling (a) or using high-resolution foveal areas (b), selectively enhancing the legibility (c).*

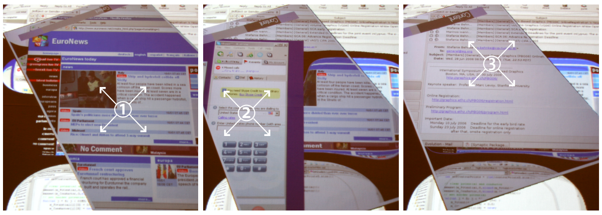

Figure 23: *The handheld projector allows for interactive navigation through the stack of overlapping display components. Individual layers are expanded and overlaid in the foveal area.*

7.2. Adaptive composite user interfaces

Figure 21 illustrates the process of creating a composite user interface. Parts of separate desktops are cut and reassembled with arbitrary scale factors. At interactive rates, the context areas of the display pieces are automatically adapted in shape to prevent collisions. In the example, a word-processor preview, the coding area of a development environment, the newest e-mails received by a mail client, the current news of a website, and the notification area of an internet phone application are all displayed side-by-side in a space-efficient manner. Our focus-and-context visualization improves the legibility of the important content, while still hinting at the peripheral information which can be explored by expanding individual displays, by scrolling the contents using gestures, or by enhancing the resolution with foveal overlays (see Figure 22). Using the portable projector and its multi-layer interaction capabilities, the user can navigate through the stack of overlapping windows (see Figure 23).

7.3. Limitations

The proposed calibration procedure of the pocket light metaphor does not support curved surfaces or multiple overlapping foveal areas. Furthermore, the mapping of the tracked frames to distinct display devices is still subject to future work. The devices could for example encode their identifications into the displayed imagery, into infrared beacons or into a physical tag attached to themselves. As a technological limitation of current handheld projectors, the image sharpness is rather sensitive to varying projection distances. Additionally, both the resolution and the brightness of today's portable projectors are still comparatively low. In the future, these problems will soon be alleviated by improved optics, better chipsets, and the potential replacement of DLP or LCD technology with a laser-based projection. Through similar improvements in camera technology, we expect to further increase the precision of our foveal tracking approach. As a minor drawback of the smooth warping of the adaptive composite interfaces, the space between the different focus areas is typically not entirely covered by content. To mitigate this problem, the users can at any time dynamically modify the degree of smoothness to fill the entire space if desired. Finally, the proposed interaction based on a thin laser layer is restricted by design to piecewise planar surfaces. Shadowing effects potentially occurring in multi-touch environments can either be accounted for in software, or be reduced with multiple overlapping light planes emanating from different directions.

8. Conclusions and future work

In this paper we have presented a set of novel techniques and metaphors for flexible and space-efficient visual workspaces which support natural and intuitive interaction. Our three main contributions include:

- A novel pocket light metaphor introducing dynamic foveal areas into visual workspaces to locally increase the resolution, provide interactive visualization and achieve content personalization. Contrary to previous work, our approach supports hybrid displays and offers a high calibration accuracy without relying on physical markers or complex hardware. We also provide a maximum degree of flexibility by not only continuously allowing for changes in size, position and resolution, but by also enabling a set of projector-based interaction modes.
- A compositing approach to tile and arrange displays in a space-saving and resolution-enhancing manner. To the best of our knowledge, our approach is the first to rely on a focus-and-context warping technique which supports

freeform cuts and combines the user-defined focus with a surrounding context area based on an adaptive reshaping.

- A unified tracking method allowing for natural interaction through bare hands, pointers and pens. It supports a single-handed and bimanual input for multiple simultaneous users and enables interaction on large ordinary surfaces. Our novel technique can serve as a robust, cost-effective and easy-to-deploy replacement for previous input methods like frustrated total internal reflection [Han05]. For a mid-air interaction, we propose a stable tracking of bare fingers based on morphological filters.

In the future, besides addressing the issues presented in Section 7.3, we would also like to extend our foveal overlays to settings where we augment not only large-scale projections, but also printed documents, display panels and posters.

Acknowledgements

We thank R. Ziegler and D. Y. Kwon for their project feedback and J. Wolf for implementing parts of the interaction.

References

[AR05] Ashdown M., Robinson P.: Experience with the Escritoire: A personal projected display. *IEEE Multimedia 12*, 1 (January-March 2005), 34–42.

[BBRF05] Beardsley P., Baar J. V., Raskar R., Forlines C.: Interaction using a handheld projector. *IEEE Comput. Graph. Appl. 25*, 1 (2005), 39–43.

[BFC05] Blasko G., Feiner S., Coriand F.: Exploring interaction with a simulated wrist-worn projection display. In *Proc. of ISWC'05* (2005), pp. 2–9.

[BGS01] Baudisch P., Good N., Stewart P.: Focus plus context screens: combining display technology with visualization techniques. In *Proc. of UIST'01* (2001), pp. 31–40.

[CB06] Cao X., Balakrishnan R.: Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Proc. of UIST'06* (2006), pp. 225–234.

[CG06] Cotting D., Gross M.: Interactive Environment-Aware Display Bubbles. In *Proc. of UIST'06* (2006), pp. 245–254.

[CNGF04] Cotting D., Naef M., Gross M., Fuchs H.: Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display. In *Proc. of ISMAR'04* (2004), pp. 100–109.

[CWDG07] Cotting D., Waschbuesch M., Duller M., Gross M.: WinSGL: Synchronizing Displays in Parallel Graphics using Cost-Effective Software Genlocking. *Parallel Computing* (2007), to appear.

[CZGF05] Cotting D., Ziegler R., Gross M., Fuchs H.: Adaptive Instant Displays: Continuously Calibrated Projections Using Per-Pixel Light Control. In *Proc. of Eurographics 2005* (2005), pp. 705–714.

[DL01] Dietz P., Leigh D.: DiamondTouch: a multi-user touch technology. In *Proc. of UIST'01* (2001), pp. 219–226.

[DP73] Douglas D., Peucker T.: Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. *The Canadian Cartographer 10* (1973), 112–122.

[FS05] Forlines C., Shen C.: DTLens: multi-user tabletop spatial data exploration. In *Proc. of UIST'05* (2005), pp. 119–122.

[Han05] Han J. Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In *Proc. of UIST'05* (2005), pp. 115–118.

[HBL*06] Haller M., Brandl P., Leithinger D., Leitner J., Seifried T., Billinghurst M.: Shared Design Space: Sketching Ideas Using Digital Pens and a Large Augmented Tabletop Setup. In *ICAT* (2006), pp. 185–196.

[HYC*98] Hung Y., Yang Y., Chen Y., Hsieh I., Fuh C.: Free-hand pointer by use of an active stereo vision system. In *Proc. of ICPR'98* (1998), pp. 1244–1246.

[Int06] Intel: OpenCV. Intel Open Source Computer Vision Library. http://www.intel.com/technology/computing/opencv/index.htm, 2006. Viewed: December 7, 2006.

[KB61] Kalman R., Bucy R.: New results in linear filtering and prediction theory. *Transactions of the ASME - Journal of Basic Engineering 83* (1961), 95 – 107.

[KIN*05] Kakehi Y., Iida M., Naemura T., Shirai Y., Matsushita M., Ohguro T.: Lumisight Table: An Interactive View-Dependent Tabletop Display. *IEEE CG&A 25*, 1 (Jan./Feb. 2005), 48–53.

[MITS05] Miyahara K., Inoue H., Tsunesada Y., Sugimoto M.: Intuitive manipulation techniques for projected displays of mobile devices. In *Proc. of CHI'05* (2005), pp. 1657–1660.

[Pin01] Pinhanez C.: Using a Steerable Projector and a Camera to Transform Surfaces into Interactive Displays. In *Proc. of ACM CHI'01* (2001), pp. 369–370.

[RBvB*04] Raskar R., Beardsley P., van Baar J., Wang Y., Dietz P., Lee J., Leigh D., Willwacher T.: RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors. *ACM Trans. Graph. 23*, 3 (2004), 406–415.

[Rek02] Rekimoto J.: SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *Proc. of CHI'02* (2002), pp. 113–120.

[RM97] Rekimoto J., Matsushita N.: Perceptual Surfaces: Towards a Human and Object Sensitive Interactive Display. In *Proc. of PUI'97* (1997).

[RS99] Rekimoto J., Saitoh M.: Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments. In *Proc. of CHI'99* (1999), pp. 378–385.

[SAKH06] Staadt O. G., Ahlborn B. A., Kreylos O., Hamann B.: A foveal inset for large display environments. In *Proc. of VRCIA'06* (2006).

[SCPR06] Stuerzlinger W., Chapuis O., Phillips D., Roussel N.: User interface facades: towards fully adaptable user interfaces. In *Proc. of UIST'06* (2006), pp. 309–318.

[TMC04] Tan D. S., Meyers B., Czerwinski M.: WinCuts: manipulating arbitrary window regions for more effective use of screen space. In *Proc. of CHI'04* (2004), pp. 1525–1528.

[VLS02] Vernier F., Lesh N., Shen C.: Visualization techniques for circular tabletop interfaces. In *Proc. Advanced Visual Interfaces 2002* (2002).

[VO90] Vuylsteke P., Oosterlinck A.: Range Image Acquisition with a Single Binary-Encoded Light Pattern. *IEEE TPAMI 12*, 2 (Feb 1990), 148–164.

[Wel93] Wellner P.: Interacting with paper on the DigitalDesk. *Communications of the ACM 36*, 7 (1993), 86–97.