Managing a Document-Based Information Space

Matthias Deller
DFKI GmbH
Kaiserslautern, Germany
matthias.deller@dfki.de

Stefan Agne
DFKI GmbH
Kaiserslautern, Germany
stefan.agne@dfki.de

Achim Ebert
DFKI GmbH
Kaiserslautern, Germany
achim.ebert@dfki.de

Andreas Dengel
DFKI GmbH
Kaiserslautern, Germany
andreas.dengel@dfki.de

Hans Hagen
DFKI GmbH
Kaiserslautern, Germany
hans.hagen@dfki.de

Bertin Klein
DFKI GmbH
Kaiserslautern, Germany
bertin.klein@dfki.de

Michael Bender
DFKI GmbH
Kaiserslautern, Germany
michael.bender@dfki.de

Tony Bernardin
University of California
Davis, CA, U.S.A
tbernardin@ucdavis.edu

Bernd Hamann
University of California
Davis, CA, U.S.A
hamann@ucdavis.edu

ABSTRACT

We present a novel user interface in the form of a complementary virtual environment for managing personal document archives, i.e., for document filing and retrieval. Our implementation of a spatial medium for document interaction, exploratory search and active navigation plays to the strengths of human visual information processing and further stimulates it.

Our system provides a high degree of immersion so that the user readily forgets the artificiality of our environment. Three well-integrated features support this immersion: first, we enable users to interact more naturally through gestures and postures (the system can be taught custom ones); second, we exploit 3D display technology; and third, we allow users to manage arrangements (manually edited structures, as well as computer-generated semantic structures). Our ongoing evaluation indicates that even non-expert users can efficiently work with the information in a document collection and that the process can actually be enjoyable.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Human Factors, Design.

Keywords: Multimodal Interaction, Interactive Search, Human-centered Design, Immersion, 3D User Interface.

IUI'08, January 13–16, 2008, Maspalomas, Gran Canaria, Spain.

INTRODUCTION

Visual information naturally and promptly evokes patterns of understanding as well as spatial associative memory. Visual processing and visual association are therefore important capacities in human communication and intellectual practice. In particular, this holds true for the more and more efficiency-driven work strategies of people in office environments. As an example, documents are not simply a means of raw information storage but rather an intricate communication instrument which has been adapted to human perception over the centuries: the choice of information carrying elements (headlines, paragraphs, graphics, images, icons, etc.), reading order, and presentation are combined so as to express the intentions of a documents author. Different combinations lead to individual document classes, such as business letters, newspapers or scientific papers. Thus, it is not only the text which captures the message of a document but also the inherent meaning of the layout and the logical structure.

Over the last decades the paradigm of document management and storage has quickly evolved towards electronic formats, leading to new means for non-tangible document processing and virtual instead of physical storage. As a result, the spatial cues/affordances of document filing and storage may be lost or abstracted. For instance, when documents are stored on portable electronic media, such as a USB stick, they can easily be taken home, but they are no longer individually tangible. To work with them they have to be retrieved from the device. The only support here are ones own memory or search engines, the latter only providing the user ”keyhole” access to the contents, ignoring most clues from the document structure and arrangement. In order to address this issue, we need to develop virtual work-environments that not only take advantage of people's inherent visual information processing capabilities but also actively motivate the user to make use of these capabilities. One of the key capabilities to exploit in virtual environments is immersion, which integrates the user more intuitively in the computer system. The user gets the impression of being part of the virtual environment and, ideally, can manipulate it similarly as one would do this in real surroundings, without devoting conscious attention to the use of an interface. For instance, 3D applications can be controlled through clever use of mouse and keyboard combinations, but the task of moving and placing objects in a 3D space with 2D interaction devices is cumbersome, requires effort and the conscious attention of the user. More complex tasks, e.g., opening and leafing through a document, require more complex combinations that users might have to initially train. One reason why virtual environments are not yet as common as one would assume given the advantages they present might be the lack of adequate (cost-effective, user-friendly, reliable, etc.) hardware and interfaces to interact with data in immersive environments, as well as methods and paradigms to intuitively interact in 3D settings.

We describe a novel perceptual system and user interface as a complementary virtual working environment for document filing and retrieval. The system:

1. provides a powerful knowledge management back-end exploiting both statistical knowledge (e.g. similarity, clustering) and formal knowledge (e.g. ontologies, stacks with labels) to assist users in relating documents.

2. presents a virtual reality implementation of the desktop metaphor supported by cost-effective 3D technology for the display (stereoscopic screens) as well as the input (six degrees-of-freedom data gloves).

3. supports functional processing of document spaces through (a) custom arrangements, searches, notes and associations; and (b) use of natural hand gestures for the navigation / manipulation of the data.

STATE OF THE ART

The problem of document space processing has been addressed from different perspectives. Welch et al. [24] proposed new desktops enabling people to ”spread” papers out in order to look at them spatially. High-resolution projected imagery can be used as a ubiquitous aid to display documents not only on a desk but also on walls or even on the floor. People at remote sites would be able to collaborate on 3D displayed objects in which graphics and text can be projected. Krohn [11] developed a method to structure and visualize large information collections allowing the user to quickly recognize whether the found information meets expectations or not. Furthermore, the user can provide feedback through graphical modification of the query. Shaw et al. [18] describe an immersive 3D volumetric information visualization system for the management and analysis of document corpora. Based on glyph-based volume rendering, the system enables the 3D visualization of information attributes and complex relationships. Two-handed interaction is made possible via magnetic trackers and stereoscopic viewing provides a user with 3D perception of the information space. The following two sections treat in more detail the two core aspects of visualization and interaction in 3D environments.

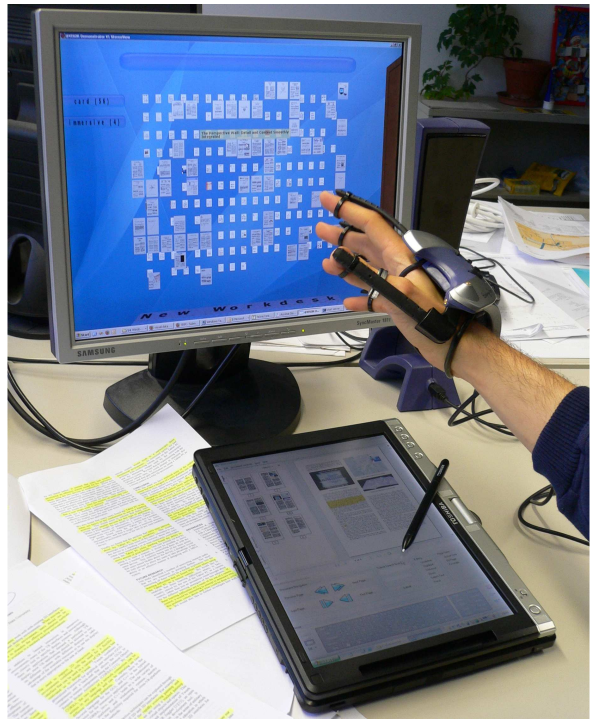

Figure 1: Our system setup: stereoscopic display, tablet pc, and data glove.

Visualization

The information cube introduced by Rekimoto et al. [14] can be used to visualize a file system hierarchy. The nested box metaphor is a natural way of representing containment. A problem with this approach is the difficulty in gaining a global overview of the structure, since boxes contained in more than three parent boxes or placed behind boxes of the same tree level are hard to perceive. Card et al. [6] presented a hierarchical workspace called Web Forager to organize documents with different degrees of interest at different distances to the user. A drawback of the system, however, is that the user, who can execute only search queries, is not assisted by the computer in organizing the documents in space as well as in mental categories.

3D NIRVE, a 3D information visualization presented by Sebrechts et al. [17], organizes documents selected from a search query based on the categories they belong to. Nevertheless, placing the documents around a 3D sphere proved to be less intuitive than simple text output.

Robertson et al. [15] developed Task Gallery, a 3D window manager that can be considered as a direct conversion of the conventional 2D metaphors to 3D environments. The 3D space is used to attach tasks to walls and to switch between different tasks by moving them on a platform. The main advantage of this approach over the 2D windows metaphor is limited to supporting a user’s spatial memory for finding existing tasks.

The Tactile 3D system [1], a commercial 3D user interface for the exploration and organization of documents, is still in development. The file system’s tree structure is visualized in

3D space using semi-transparent spheres that represent folders and that contain documents and other folders. The attributes of documents are at least indicated through the use of different shapes and textures. Objects within a container can be placed in a sorting box and be sorted by various sorting keys in the conventional way, forming 3D configurations like a double helix, pyramid, or cylinder. The objects that are not in the sorting box can be organized manually.

Focusing on user-experience Agarawala et al. [2] propose an implementation of the virtual desktop that incorporates a simulation of the physical affordances of documents: friction and mass influence their motion when dragged or tossed, and collisions with other documents are simulated realistically. Their environment allows the user to easily pile and group documents in more casual arrangements. To support such tasks they introduce intuitive interaction metaphors, such as the Lasso-Menu that enables users to circle documents to quickly select, pile or otherwise manipulate them. While the arrangement capabilities of the environment are powerful when combined with human manipulation, contextual information, i.e., document content and meta data, is not exploited and relations between documents are only supported through their placement on the desktop.

The ongoing discussion on the usefulness of 3D visualization of complex information spaces is quite controversial [17], [23], [21], [25]. Nevertheless, Ware's results [23] show that building a mental model of general graph structures can be improved by a factor of three compared to 2D visualization. The studies also show that 3D visualization supports the use of the spatial memory and that user enjoyment is generally better with 3D visualizations, which has been recently rediscovered as a decisive factor for efficient working.

Interaction

Research on human-computer interaction has devoted much attention to visual capturing and interpretation of gestures. Either the user or just hands are captured by cameras so position or posture of the hands can be determined with appropriate methods. To achieve this goal, several different strategies have been proposed. The most natural and most comfortable way to interact uses non-invasive techniques. Here, the user is not required to wear any special equipment or clothing. Consequently, the system's software is tasked with interpreting a set of cameras' live video streams in order to identify the user's hands and the gestures made. Some approaches aim to solve this segmentation problem by assuring a specially prepared background [4], still others try to determine hand position and posture by feature recognition methods [13]. Newer approaches use a combination of these methods to enhance the segmentation process and find the user's fingers in front of varying backgrounds [22]. Other authors simplify the segmentation process by introducing restrictions, often by forcing the user to wear marked gloves [19], using specialized camera hardware, or restricting the capturing process to a single properly prepared setting [7].

Although promising, all of these approaches have the common drawback that they impose special needs on the environments in which they are used. They require uniform, steady lighting conditions, high contrast in the captured pictures,

Figure 2: A virtual work desk.

and they have difficulties when the user's motions are so fast that hands are blurred on the captured images; they demand substantial computing resources as well as special and often costly hardware; and the cameras for capturing the user have to be firmly installed and properly calibrated, restraining the user to a predefined area to allow gesture recognition. Often, a separate room has to be used to enable the recognition of the user's gestures. Alternatively, gestures can be captured via special interface devices such as data gloves [20, 10].

VISUALIZATION OF DOCUMENT-BASED INFORMATION SPACES

A collection of documents can be considered as an information space. The documents as well as the relations between them carry information. This information is refined and grows as the space is processed. Information visualization techniques, combined with the ability to produce realistic real-time, interactive applications can be leveraged to create a new generation of document explorers. People should be more comfortable using an application environment that resembles a real one. Users can perceive information about the documents, such as their size, location, and relation to other documents, by using their natural capabilities to absorb spatial layouts and to navigate in 3D environments. In addition,cognitive capacities are freed by shifting part of the information-finding load to the visual system. This results in a more efficient combination of human and computer capabilities: Computers have the ability to quickly search through documents, compute similarities, generate and render document layouts, and provide other tools not available in paper document archives; humans can visually perceive irregularities and intuitively interact with 3D environments. Last but not least, a well-designed virtual reality-like graphical document explorer should be more enjoyable to the average user than a conventional one and, thus, more motivating [26].

Figure 2 shows the main area of the workspace. Here, documents can be placed, arranged in ordered or unordered stacks; reminders and input for "to-do's" can be retained; and current projects and duties can be grouped. Elements in the visible documents marked in yellow, indicate that something was attached to them, either notes, or links leading to other

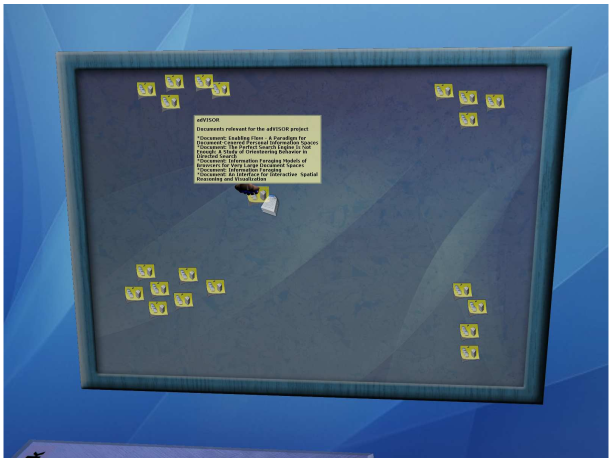

Figure 3: A virtual post-it pin board.

Figure 4: Searching for a document. The different appearances of the documents in size on the projected screenshot relate to their position closer to or farther from the user in the 3D environment.

entities including similar elements, documents, to-do's, or stacks. Apart from its function as a spatial store for document clusters, the work desk also functions as a representation of the user's current task. At any time arrangements can be saved and later restored to facilitate switching between tasks. The hand shown in Figure 2 points to the left-arrow button, which pops-up a preview of the most recently saved workspace arrangement. If the button is pushed, the current workspace is saved and replaced by the stored one.

Figure 3 depicts the pin-board, used as a "store" for post-it like notes, to-do's, and reminders. Links to documents, document stacks, and whole working desks can be easily grouped in a post-it for quick access to related material. The post-its themselves can equally be grouped and clustered on the pin board to represent mental models of the user. Also, user post-its on the pin board remain unaltered by task switches. In this manner, it is easily and intuitively possible to carry over annotated documents or document stacks from one task to another.

Visualization of Documents

It is natural to work with physical assortments of documents, ordered collections of books, more or less ordered stacks of documents, and unordered heaps of documents. Thickness, wear, front layout, pictures, headlines, scribbles (even if almost unreadable), coffee-cup marks, position and stacking help to maintain the content, context, and purpose of the documents and to support a person to do the necessary work.

There are different recurring elements in working with documents. A new document goes through different stages in which it is increasingly understood by the user. First, perhaps only the title gets consciously memorized. Next, some spotted buzzwords enrich the image of the contained information. In the end, the full message of the document is completely understood and available in one's mind, i.e., linked to concepts in the mind. Throughout this process the document needs to be put back onto the desk, perhaps onto an appropriate stack, be accidentally or consciously retrieved and receive attention (reading, reflection, association), typically more than once. Let us consider an example to illustrate how such minor details can be important: a document might be inside a stack of documents, with a corner somewhat poking out. This might prompt a user to retrieve this document at some given time in the process. Otherwise, the document might be overlooked for a longer time. The seemingly minor detail, thus, plays an important role in a larger process involving the document. The idea is to create a system that provides as many such possible details. Still, the basic approach is simple, create a virtual desktop.

It is clearly a goal to transcend the possibilities of a wooden desktop, and, for example, let the computer make documents fly, let them display links to each other, and get them arranged in one or the other order at the push of a button.

Figure 4 shows documents in the so-called document search mode. Documents are arranged soaring in a matrix. The 3D space is used to highlight documents that match keywords the user entered by having them float closer to the user, subtly and naturally drawing the attention to them. The depictions show thumbnails of the first pages and their thickness indicates the number of pages. Pointing at a document causes a label to pop up with the file name and an enlarged view of the front page in the upper-right corner of the screen. Documents can be marked to keep them in sight or to apply actions. At any time a view can be saved and restored later.

Stacks of documents can be selected and conveniently browsed similar to the functionality of Windows Vistas new application switcher, see Figure 5. At any time the lower and the upper parts of the stack (in the figure respectively the left and right illustrations) and the front page of the document at the current position in the stack (middle illustration) are visible. With a smooth sliding motion, a user can make all documents in the stack move, one by one, from the lower part to the upper part, seeing them highlighted in the center as they move. The user can stop, reverse, and continue the fly-by anytime.

With a document in focus an analogous kind of browsing is possible, depicted in Figure 6. Sliding motions make the fo-

Figure 5: Browsing a document stack.

Figure 6: Browsing a document.

cus skim through the document, from the first page to the last. At any time the pages before and after the current page in focus are visible to the left and right of it. Thus, an inherent perspective focus on the current page is provided, resembling Mackinlay et al's perspective wall [12]. In the example in Figure 6, there is a note attached to the visible page, depicted as a yellow note sheet nailed to the page. The yellow line at the bottom edge of the page reveals that the note refers not only to an element in the page but to the whole page. Also, the line continues to the previous and following pages in decreasing intensity. In this way, a user browsing the document can take note of highlighted pages in the vicinity of his currently focused page and quickly and intuitively navigate there by simply following the line in the direction of increasing intensity.

Certainly, at times the user needs to overview all pages of a document at the same time, analogously to search mode, where all documents of a collection are visible at the same time. The feature has been a part of popular software like Word and PDF readers and is also implemented in our system (see Figure 7). Up to a certain limit, thumbnails of the pages are arranged in the shown manner. Documents containing more pages than the reasonably presentable limit are distributed across several overview pages.

Visualization of Relations

Users enter a document space and work with it by discovering and maintaining relations between documents. For example, a new scientific article might be discovered relevant to a project. Thus, users need to maintain relations that they discover. In addition to user generated associations, state-of-the-art technology implemented in our system provides users with access to a semantic engine, which calculates similarity relations between documents [3].

Relations between documents can intuitively be represented by connecting the document under the pointer to the related ones using red laser beams as shown in Figure 8. These beams create a mental model like that of thought flashes moving the user's attention from the current document to related documents. Additionally, semantic links between documents can be visualized in variable intensity, giving the user a quick overview of the most similar documents. The advantage of this way of representing relations is that documents are strongly connected visually by curves which create additional spatial patterns that stimulate the user's visual processing.

Unfortunately, the thought flashes can also occlude documents. To overcome this issue, documents which are related to the document pointed to can simply be highlighted as shown in Figure 9. In our case, they are colored in red. In this way the document body is not occluded by the visualization of the relationships. Also, the highlighting can be directly used in *document search mode*.

INTERACTING WITH DOCUMENTS

The most natural way for humans to manipulate their surrounding, including documents on their desktop, is to use their hands. Hands are naturally used to grab, move, point at, mark, and manipulate objects of interest. Also, in order to communicate intentions, hands can indicate postures and gestures. This is done without consciously having to think about it, and without interrupting current tasks. Therefore, we consider a gesture recognition engine to minimize the cognitive load required for learning and using a user interface in a virtual anvironment.

The gesture recognition system should be adaptable to various conditions, like alternating users or hardware (possibly even mobile devices.) It should also be fast and powerful enough to enable a reliable recognition of a variety of gestures without hampering the performance of the actual application.

Hardware Setup

The data glove we used is a P5 from Essential reality, originally designed for gaming, see Figure 1. It features five sensors for the fingers, and an infrared tracking system for the glove's position and orientation. The tracking also requires a base station (visible in the background of the picture) with infrared sensors. The P5 costs about 50 Euro, which is cheap when compared to the cost of about 4000 Euro of typical professional data gloves. The measurements of the flexion of the

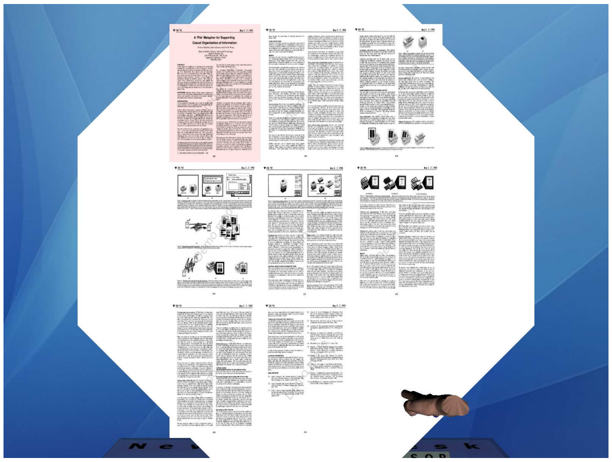

Figure 7: Document overview mode.

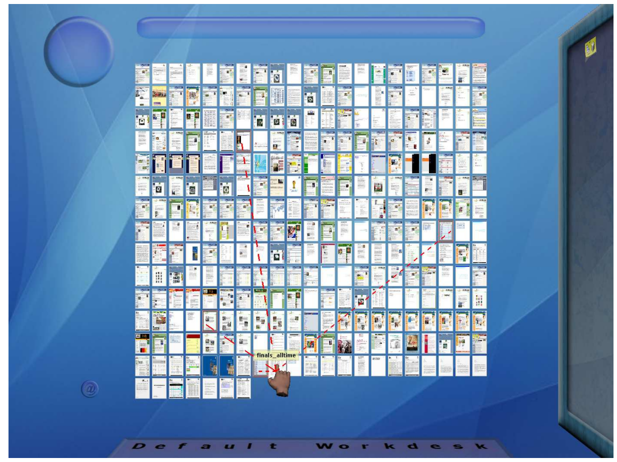

Figure 8: Highlighting relations between documents.

fingers are quite accurate. Also, the estimated position information is quite dependable. However, the measurements of yaw, pitch and roll of the glove are, depending on lighting conditions, very unreliable with sudden jumps in the data. We had to develop additional filtering mechanisms to acquire sufficiently reliable values.

Posture and Gesture Recognition and Learning

A major problem for the recognition of gestures by visual tracking is the high amount of computational power required to determine in real time the most likely gesture carried out by the user. Rendering a virtual environment concurrently can hardly be done on a single average consumer PC. Our reliable real-time recognition is capable of running on any current workplace PC and can easily be integrated in normal applications without monopolizing much of the system's processing resources. Like Bimbers "fuzzy logic approach" [5], we use a set of gestures that have been previously taught by performing them in order to determine the most likely match. However, for our system we do not define gestures as motion over a certain period of time, but as a sequence of postures made at specific positions with specific orientations of the user's hand.

The postures are composed of the flexion measurements of the fingers, the orientation data of the hand, and a value indicating the relevance of the orientation for the posture. For a pointing gesture with stretched index finger, the orientation and position of the hand may be required to determine what the user is pointing at, but the gesture itself is the same, whether he is pointing at something to his near left or his far right. For some gestures the orientation data is much more relevant. For example, the meaning of a fist with the thumb stretched out can differ significantly depending on the thumb pointing upward or downward. In other cases, the importance of orientation data can vary. For instance, a gesture for dropping an object may require the user to open his hand with the palm pointing downwards, but it is not necessary to hold his hand exactly leveled. It is easy to teach the system new postures that may be required for specific applications. The user performs the posture, captures the posture data by hitting a key, names it and sets its orientation quota.

Alternately, existing postures can be adapted for specific users. To do so, the posture in question is selected and performed several times by the user. The system captures the different variations of the posture and determines a resulting averaged posture definition. In this manner, it is possible to create a collection of different postures, a so-called posture library. This library can be saved and loaded in form of a gesture definition file, making it possible for the same application to have different posture definitions for different users, which can even be changed on-the-fly.

Data Acquisition and Filtering

The recognition module consists of two components: the data acquisition and the gesture manager. The data acquisition pipes the tracking data through several filters. Changes in the position or orientation data that exceed a given deadband limit are discarded and replaced with their previous values to eliminate changes in position and orientation that are most likely erroneous. The resulting data is then straightened out by a dynamically adjusting average filter. Depending on the variation of the acquired data, the size of the filter is adapted within a defined range: if the data is fluctuating in a small region, the size of the filter is increased to compensate jittered data. If the values show larger changes, the filter size is decreased to reduce latency in the acquired position and orientation. The resulting data are good enough for the matching process of the gesture manager.

The gesture manager compares the data to the known postures. It identifies a matching posture if it is held for an adjustable minimum time span. In tests we found that values between 300 and 800 milliseconds are suitable to allow for a reliable recognition without forcing the user to hold the posture for too long. Every recognized posture is sent to the application that started the acquisition thread, accompanied by: a time stamp, the string identifier of the recognized posture, the previous posture, and the position and orientation of the glove at the moment of the posture.

Beyond postures, our system keeps track of movements and buttons. Greater changes of position or orientation fire Glove-Move events (data-wise comprising start and end values of

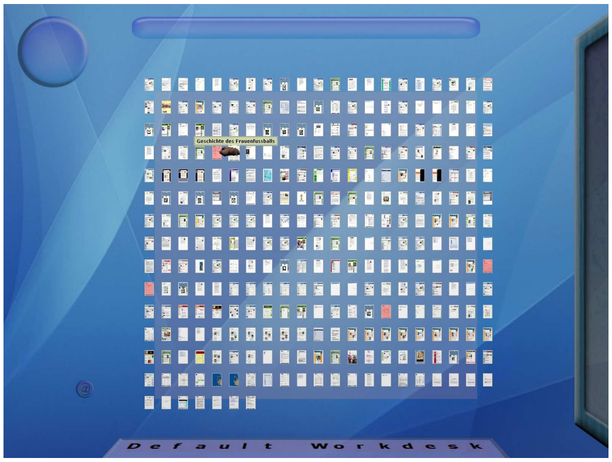

Figure 9: Highlighting relations by coloring.

position and orientation); buttons like those of the P5 cause ButtonPressed and ButtonReleased events.

The data acquisition process can easily be adapted to any other data glove and more kinds of devices, either for mere posture recognition or in combination with any additional six-degrees-of-freedom tracking device like the Ascension Flock of Birds to achieve full gestural interaction.

Gesture Management and Recognition

The gesture manager maintains the list of known postures and provides functions to manage the posture library. As soon as the first posture is added to the library or an existing library is loaded, the gesture manager begins matching the data received from the data acquisition thread to the stored data sets. This is done by first looking for the best matching finger constellation. In this first step, the bend values of the fingers are interpreted as five-dimensional vectors, and for each posture definition the distance to the current data is calculated. If this distance fails to be within an adjustable minimum recognition distance, the posture is discarded as a likely candidate. If a posture sufficiently matches the data, the orientation data is compared in a likewise manner to the measured values. Depending on whether this distance exceeds another adjustable threshold, the likelihood of a match is lowered or raised according to the orientation quota associated with the corresponding posture data set. This procedure is very reliable, supports fast matching of postures, and makes possible a consistent recognition.

In addition to determining the most probable posture, the gesture manager provides several means to modulate parameters at run time. New postures can be added, existing postures adapted, or new posture libraries loaded. The recognition thresholds can be adjusted on the fly, such that it is possible to start with a wide recognition range to enable correct recognition of user postures without posture definitions adapted to a specific person. As the postures are customized through use, the boundaries can be readjusted to more appropriately constrain matches.

The recognition of single postures like letter-postures of the American Sign Language ASL is as easily done as the recognition of more complex, dynamic gestures. Gestures which are sequences of successive postures are recognized by tracking the sequence of performed postures as a Finite State Machine. In this manner, almost any desired gesture can quickly be implemented and recognized.

Dimensional congruence and natural interaction

While interacting in virtual environments, situations that require 2D input, 2D output, or both will occur commonly. A 3D environment has its merits when working with multiple documents at the same time, for ordering, arranging and sorting documents. Other interactions, however, like reading or editing a single document, are two-dimensional operations. The use of an additional dimension in this case is not only unnecessary, but may actually reduce the readability of the document.

It is necessary to match the dimensionality of interaction with the dimensional demands in order not to sacrifice task performance. This is referred to as *dimensional congruence*, a term coined by Darken and Durost [8]. Obviously, it is difficult to determine a best interaction technique or class of interaction techniques for performing a certain task. But, 3D tasks are best executed by 3D techniques, and 2D tasks are best executed by 2D techniques. In order to construct interfaces that can handle both equally well, an adequately priorized combination needs to be considered. For practical reasons, a designer might decide to sacrifice performance on those 2D tasks in favor of dominant 3D tasks. The authors demonstrated that 3D interaction techniques were preferable on 3D dominant tasks, purely 2D interaction techniques were preferable on 2D dominant tasks, and their hybrid interface where 2D and 3D interaction techniques were matched to each individual task showed the best performance.

After several user studies with earlier demonstrators [9], we designed and implemented a dimensional congruent prototype for visualization of and interaction with personal document spaces. We allocated the different visualization and interaction metaphors to a 2D + 3D display environment by matching task, device, and interaction technique. To achieve this, we used an auto-stereoscopic display in combination with an optically tracked P5 data glove for 3D visualization and interaction. Additionally, a Toshiba Tablet PC for higher-resolution 2D visualization and pen interaction was placed horizontally in front of the stereo-display.

EXPERIMENTS

We have evaluated our system for several virtual document spaces, in which the users could manipulate documents and trigger actions by performing gestures. For a strong immersive experience, we used the demonstration setup shown in Figure 1, including a 3D view provided by a stereoscopic display, the SeeReal C-I. This monitor creates a real three-dimensional impression of the scene by showing one perspective view for each eye and separating them in a prism layer on the screen itself. No glasses are required. As a compromise, the achieved resolution of the image is lower than on regular displays, the effect of which are especially obvious while displaying texts. To take this into account we implemented our prototype as a dimensionally congruent system by adding a Toshiba Tablet PC for 2D interaction tasks.

The test subjects were given a video introduction of the system, then they had time to experiment with the demonstrator for themselves and to get used to the data glove. We used two different kinds of gestures, semiotic and ergotic gestures [16]. Semiotic gestures are used to communicate information (in this case pointing out objects to mark them for manipulation); ergotic gestures are used to manipulate a person's surroundings. In the demonstrator, the test subjects could grab objects, to drag them to another position in the virtual environment, or to activate them (for example open a document). In this case, the system wasn't trained for each user, but used a generic set of gesture definitions (point, grab, drop, browse left/right, and open/activate) for all users. Nevertheless, after about ten minutes of using the glove, users were able to manipulate the virtual environment to their satisfaction. Because of the imprecise glove hardware and since the focus of the experiment was not on the gesture recognition, we didn't measure exact error rates, but rather asked the subjects for an informal appraisal of their satisfaction with the glove interaction. Our demo scenario presented a virtual desk, with different randomly arranged documents. A pin-board and a calendar were visible in the background. An avatar of the user's hand was also rendered in the scene, so as to provide feedback on how the system reacted to his moves.

The principle part of the demonstration scenario running on our framework takes place in a virtual work environment on the stereoscopic display. The virtual workspace consists of the desktop in the lower visual region of the displayed environment, and the virtual pin board to the right of the virtual viewpoint. When interacting with each of these objects, the virtual camera is moved from the original perspective to provide a full view on the table surface or pin board. The desktop provides a surface to arrange documents and document clusters spatially, while additionally representing the current task of the user, explained in detail later. Normally, most of the visible region of the stereoscopic screen is filled by a view on the cover pages of all available electronic documents. Initially, all documents hover in space at the same distance from the user. He can interact with them in different ways using the data glove. Each interactive part of the virtual environment is tagged with three-dimensional tool tips, hovering directly in front of the currently focused object and providing a short summary of the object, e.g. document title and number of pages.

He can grab documents and drag a copy of the document to the desk surface, arranging them in his own meaningful fashion. He can also put documents on top of each other, thereby manually creating document clusters which are in turn represented as stacks on the surface. He can also rearrange existing stacks, combine them and name them, causing a name label to appear directly above the corresponding stack. Also, by dragging either a document or a stack to the virtual pin board, he can create an intelligent post-it. While doing so, he can provide a title and additional text, causing a post-it representation to appear both on the pin board and on the linked document or stack. Additional documents and stacks can be added to or removed from the post-it by simply dragging the corresponding object onto it. Furthermore, the user can drag post-its onto the table surface, thereby creating a new stack

Figure 10: Virtual desktop demo.

containing all documents and stacks linked with it. In this way, post-its provide an additional way to cluster documents, independent of the current work desk and therefore independent of the user's current task.

Of course, there is also the need to provide views inside a document stack or to view a single document. Content of a selected cluster is examined by performing a "grab and open" gesture on the stack, causing the stack to move to the left side of the screen and the rest of the environment to blend out. The one-dimensional interaction of browsing through the stack is done by using the Tablet PC, either by clicking on a direction arrow or by dragging the pen in the corresponding direction. While browsing, the documents on the stack fly first toward the user to provide a detailed view of the document, then to a second stack to the right, representing browsed documents. The user can also browse individual documents, either starting from the original document representation or by choosing a document from a document stack. Because of the higher resolution of the 2D screen and because 3D is more of a hindrance when reading two-dimensional text, browsing as well as editing of documents can be done completely on the Tablet PC. After completion, the visualization changes back to the three-dimensional representation.

In addition to the proposed natural interaction metaphor, the user can also use the benefits of electronic documents, i.e. he is able to search through documents as well as attached annotations. This is done by activating a search bar positioned above the floating document representations. The user can then enter a sequence of search terms on the Tablet PC, causing individual documents to leave the original document plane and move towards or away from the user, depending on how good his search queries are matched. In this manner the user gets a pre-attentive clue on which documents best match his current search, as well as providing a natural zooming on important documents. Entered search terms are represented as hovering text fields to the left of the user. By dragging one or more of these to the work desk surface, he can create new stacks containing only the matches for corresponding search terms. Moreover, he can manually redefine the importance of documents by touching one of the representations and pulling it towards him or pushing it away. By doing so, documents that the information back-end has identified as having similar content are also rearranged to represent the user's focus of interest. More abstract interaction possibilities like changing the search mode or displaying all Post-Its attached to a single document can be invoked by a three-dimensional ring menu. A pointing gesture at a virtual object triggers a ring menu to appear at the object's spatial position. By moving his hand up, down, left or right, the user can select entries of the according menu.

Finally, the user is able to switch between different tasks, e.g. from searching documents for writing a paper to assembling documents for a meeting. As mentioned before, these tasks are represented by the work desk. To change his active task, the user can activate either a 3D widget in front of the virtual work desk or on the 2D GUI on the Tablet PC. When a new task is created, the current work desk is saved, complete with all stacks on the work desk, current search terms and arrangement of document representations. A new empty work desk is created and the search as well as document representations are reset to the initial state. In this way, the user can quickly switch between tasks without destroying the context he has previously created.

We had several user-groups (knowledge workers, secretarial assistants, students, and pupils) test our environment. They were given specific tasks that simulated realistic document space processing and stressed the capabilities of our system. In particular, users were asked to perform several searches within documents and over the document body; produce corresponding arrangements; interrupt tasks to address new ones; resume interrupted tasks and link their findings. This was presented in the context of a journalist reporting on the end-game of a soccer world cup. After that, the users had to repeat the same tasks using conventional search engines and file systems.

A broad majority of the users gave encouraging feed-back: 74% accepted the visualization and navigation metaphors. The spatial information item layout was reported to be intuitive by nearly 90%. The movement of important documents to the front was selected as the best choice by 72% even before they saw our interface. As there is still no commonly accepted standard for interaction with 3D environments it took the users some time to understand what kind of interaction possibilities they actually have. When they got used to the system they often instantly came up with proposals for features that could be added. We understand this to be due to the immersive and intuitive effect of 3D interfaces as well as their enormous creative potential.

CONCLUSIONS AND FUTURE WORK

Human thinking and knowledge work are heavily dependent on sensing the outside world. One important part of this perception oriented sensing is the human visual system. It is well-known that our visual knowledge disclosure i.e., our ability to think, abstract, remember, and understand visually - and our skills to visually organize are extremely powerful. Our overall vision is to realize a customizable virtual world which inspires the users thinking, enables the economical usage of his perceptual power, and considers a multiplicity of personal details with respect to thought process and knowledge work. We have made major steps forward towards this vision, created the necessary framework, and developed needed modules.

We conclude that by creating a framework that emphasizes the strengths of both humans and machines in an immersive virtual environment, we can achieve great improvements in the effectiveness of knowledge workers and analysts. We strive to complete our vision by further extending our methods to present and visualize data in a way that integrates the user into artificial surroundings seamlessly and provides the opportunity to interact with it in a natural way. A holistic context and content-sensitive approach for information retrieval, visualization, and navigation in manipulative virtual environments was introduced. We addressed this promising and comprehensive vision of a more efficient man-machine interaction in manipulative virtual environments by "immersion": a frictionless sequence of operations and a smooth operational flow, integrated with multi-sensory interaction possibilities, which supports interaction of human work activities and machine support. Ultimately, this approach will enable a powerful immersion experience: the user will have the illusion that he is actually situated in the artificial surroundings; the barrier between human activities and technology will vanish; and the communication with the artificial environment will be seamless and homogeneous. Visually driven thinking, understanding, and organizing will be promoted, and the identification and recognition of new relations and knowledge facilitated.

We have dedicated our studies on virtual environments to personal information spaces which are, to a high degree, based on documents, i.e., personal document-based information spaces. Our usability tests have produced promising results for highly developed visualization and interaction metaphors. It is natural and easy to handle documents, and to obtain new ideas from computer-generated clues (like similarity relations and clusters) about documents.

As a next step, we plan to transfer our insights about personal information spaces, suited to one user, to a multi-user environment, using a large, stereoscopic display device: the PowerWall. This wall-sized screen facilitates a whole new area of visual and interaction metaphors for virtual information spaces. Because of the much larger screen real estate, users often can't see the whole screen at once. Also, users are able to move in front of the screen, requiring adequate techniques for focus and context visualizations, as well as interaction paradigms that consider the position and movement of users.

ACKNOWLEDGMENTS

This research is supported by the German Federal Ministry of Education and Research (BMBF) and part of the project @VISOR. Is also supported by DFG's International Research Training Group "Visualization of Large and Unstructured Data Sets".

REFERENCES

1. Tactile: Tactile 3d user interface. http://www.tactile3d.com/, 2007.

2. A. Agarawala and R. Balakrishnan. Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2006.

3. Stefan Agne, Christian Reuschling, and Andreas Dengel. Dynaq - dynamic queries for electronic document management. In *EDOC Workshops*, page 61. IEEE Computer Society, 2006.

4. G. Appenzeller, J. Lee, and H. Hashimoto. Building topological maps by looking at people: An example of cooperation between intelligent spaces and robots. In *Proceedings of the IEEE-RSJ International Conference on Intelligent Robots and Systems*, 1997.

5. O. Bimber. Continuous 6dof gesture recognition: A fuzzy-logic approach. In *Proceedings of 7-th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media (WSCG'99)*, 1999.

6. S. Card and Robertson. G. The webbook and the webforager: An information workspace for the world wide web. In *Proceedings of the Conference on Human Factors in Computing Systems*. CHI'96, 1996.

7. J. Crowley, F. Brard, and J. Coutaz. Finger tracking as an input device for augmented reality. In *Automatic Face and Gesture Recognition, Zrich*, 1995.

8. Rudolph P. Darken and Richard Durost. Mixed-dimension interaction in virtual environments. In *VRST '05: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 38–45, New York, NY, USA, 2005. ACM.

9. Andreas Dengel, Stefan Agne, Bertin Klein, Achim Ebert, and Matthias Deller. Human-centered interaction with documents. In *HCM '06: Proceedings of the 1st ACM international workshop on Human-centered multimedia*, pages 35–44, New York, NY, USA, 2006. ACM Press.

10. T.S. Huang and V.I. Pavlovic. Hand gesture modeling, analysis, and systhesis. In *Proc. of International Workshop on Automatic Face- and Gesture-Recognition (IWAFGR), Zurich, Switzerland*, 1995.

11. U. Krohn. Visualization for retrieval of scientific and technical information. Dissertation, Techn. Univ. Clausthal,, 1996. ISBN 3-931986-29-2.

12. Jock D. Mackinlay, George G. Robertson, and Stuart K. Card. The perspective wall: detail and context smoothly integrated. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 173–176, New York, NY, USA, 1991. ACM.

13. J. Rehg and T. Knade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *Proc. ECCV*, volume 2, 1994.

14. J. Rekimoto and M. Green. The information cube: Using transparency in 3d information visualization. In *Workshop on Information Technologies & Systems*, pages 125–132, 1993.

15. G. Robertson and M. van Dantzich. The task gallery: A 3d window manager. In *CHI Conference*, pages 494–501, 2000.

16. A. Sandberg. Gesture recognition using neural networks. Masterthesis, 1997.

17. M. Sebrechts and J. Cugini. Visualization of search results: A comparative evaluation of text, 2d, and 3d interfaces. *In Research and Development in Information Retrieval*, pages 3–10, 1999.

18. C.D. Shaw, J.M. Kukla, I. Soboro, C.K. Ebert, D.S.and Nicholas, A. Zwa, E.L. Miller, and D.A. Roberts. Interactive volumetric information visualization for document corpus management. *International Journal on Digital Libraries*, pages 144–156, 1999.

19. T. Starner, J. Weaver, and A. Pentland. A wearable computer based american sign language recognizer. In *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998.

20. T. Takahashi and F. Kishino. Hand gesture coding based on experiments using a hand gesture interface device. In *SIGCHI Bulletin*, 1991.

21. Frank van Ham and Jarke J. van Wijk. Beamtrees: Compact visualization of large hierarchies. In *INFOVIS '02: Proceedings of the IEEE Symposium on Information Visualization (InfoVis'02)*, page 93, Washington, DC, USA, 2002. IEEE Computer Society.

22. C. Von Hardenberg and F Brard. Bare-hand human-computer interaction. In *Proceedings of the ACM Workshop on Perceptive User Interfaces, Orlando, Florida*, 2001.

23. C. Ware and G. Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. In *ACM Transactions on Graphics 15*, volume 2, pages 121–140, 1996.

24. G. Welch, H. Fuchs, R. Raskar, H. Towles, and M.S. Brown. Projected imagery in your "office of the future. *IEEE Computer Graphics and Applications*, 20(4):62–67, Jul/Aug 2000.

25. U. Wiss and D. Carr. An empirical study of task support in 3d information visualizations. In *IV Conference Proceedings*, pages 392–399, 1999.

26. Ping Zhang and Na Li. The importance of affective quality. *Commun. ACM*, 48(9):105–108, 2005.