Experiments towards 3D Immersive Interaction for Digital Libraries

Rodrigo Almeida, Pierre Cubaud, Jérôme Dupire, Stéphane Natkin and
Alexandre Topol

Conservatoire National des Arts & Métiers (CNAM)
Centre d'Etudes et de Recherche en Informatique (CEDRIC)
292 rue St-Martin, 75003 Paris, France

Abstract. We describe in this paper an experimental setting for browsing and reading simultaneously multiple digital documents. An hemispherical visualization device is used to immerge the reader into 3D representations of digital collections.

1   Introduction

Relying on digital resources for academic researches is now a reality for many scientists and students, but it is usually restricted to journals, proceedings or bibliographic databases. For humanistic studies, a step forward is the digitalization of the historical sources themselves. Such documents are usually rare and fragile, so they are kept in dedicated storerooms within libraries. The digitization of these documents is an appealing answer to the problems of preservation and accessibility. Digital library (DL) technologies have benefited from the impressive increase of digital data capture, storage and transmission capabilities along with the consequent fall of their cost [1, 2].

Because of the multiplication of digital libraries, users must now deal with a new problem: evaluating the adequacy of these collections with their documentation needs. To estimate the value of a corpus, users usually submit queries to a search engine. This implies of course the user's capability to enunciate such a query. How therefore is it possible to stroll in a digital library ? On the other hand, what visualization techniques should be used when the size of the digital holdings may reach a million items ? DL also suffer from the limitation of the WIMP (windows, icons, menus, pointers) techniques which form the basis of today's interfaces. The consultation of multiple documents in parallel is limited by windows overlapping and their constant need for resizing (fig. 1). As a consequence, most DL users prefer documents downloading and printing. In this respect, Web-based interfaces do not provide much more comfort than other, much older, reading technologies.

In 1588, the italian engineer Agostino Ramelli published in France his Various and Ingenious Machines [3]. This book deals mainly with pumping and milling techniques, but one of the 195 mechanisms described is particularly interesting for the history of information visualization. Ramelli's Reading Wheel (fig. 2 left) enables its user to switch rapidly and effortlessly from one given book to another.

The wheel could be built with arbitrary dimensions, according to the reader's collection (two dozens books was a large figure at that time). Although it has never been built, the Reading Wheel is acknowledged as a first focus+context apparatus for books reading. The second important point with Ramelliś Wheel is the predominant use of spatial memory and gestures. Indeed, the reader sat in front of the device remembers the necessary rotation to switch from one book to another. To do such an operation, the gesture will depend upon the rotation angle needed. User's movements are therefore directly linked to the navigation through books stored on the device. Today's digital documents and libraries technologies clearly encompass the Reading Wheel storage capabilities and are not limited to one media type, such as books. One must however admit that, in daily routine work with digital documents, scholars are often frustrated by their equipment (fig. 2 right) and may regret the comfortable vizualisation and direct manipulation capabilities of the Reading Wheel.

We describe in this paper an experimental setting for an immersive and interactive display for digital documents. This work is part of an ongoing project of digitization of antique books related to french history of technology : the Conservatoire Numérique des Arts et Métiers. Along with a general public Web service (http://cnum.cnam.fr), we study how 3D interactive techniques can help to overcome the limitations of the WIMP-based Web interfaces. Our approach explores both context-view and detail-view. Multiple-document simultaneous view is essential for cross-referencing documents. In addition, focusing on a single document, reading it, and observing it thoroughly is also important in this framework. In order to concomitantly handle these two features, we conceived a model allowing for peripheral awareness. The main idea is to use a small display for a single document reading while a larger immersive display for multiple document view and contextual information. After a review of related works, we briefly present in the following chapter the 3D user interface of the reading workbench. The experimental visualization device is then discussed. A few guidelines for future work are presented as a conclusion.

2   Related work

Browsing through library collections with a virtual reality approach has been studied by many authors. 3D representations of existing libraries have first been experimented with VRML at the IRCAM Library [4]. Similar experiments are described in [5] and [6]. More metaphorical representations are needed when the books are not hold in public rooms, such as antiquarian collections. This is also the case for digitization projects that are not related to one specific collection. [7] describes a 3D interface for browsing through a large, nation-wide, repository of thesis. A step further in abstraction was made in the LibViewer project [8]. Metadata documents are queried by the user and the matching results are presented within pseudo-3D shelves.

The projects mentioned above did not deal with the reading of the documents. A complete 3D environment is described in the Web Forager project [9].
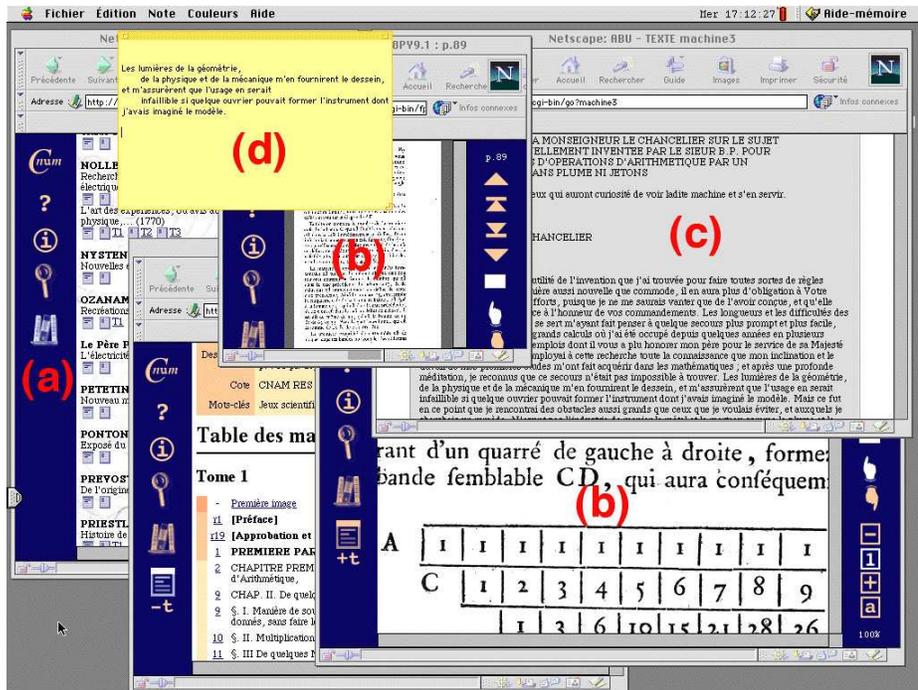
Fig. 1. Screenshot of a working session with CNAM digital libraries : (a) catalogue (b) facsimiles of book pages (c) fulltext reproduction (d) end-user document extract.
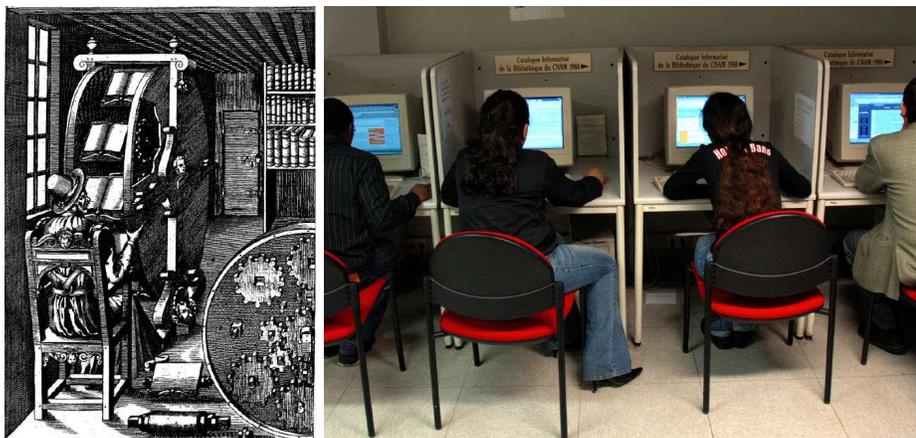
Fig. 2. The Reading Wheel of A. Ramelli (left) Working today at the CNAM library (right).

In this interface, the documents (Web books) are the main elements and arrangement tools are produced on demand as secondary pop-up resources. In our own work [10, 11], we followed a somewhat different approach. The collection is the main element of the interface since the user's interest and work depend upon it. The display of scanned pages also imposes different constraints than the HTML documents handled by the Web Forager. Realistic page turning was pioneered at the British library for a public exhibition of rare books in 1998 [12]. It is further refined in [13, 14] with full 3D navigation. In these experiments, only one book at a time can be read by the user.

Immersive visualisation enhances user experience in a multiple document browsing and reading context. Working simultaneously with multiple documents is a complex task. Standard visualization and interaction devices have been shown in [15] to be relatively inefficient when compared to paper support for reading and writing documents. [16] compares traditional computer workspaces to aeroplane tray tables. In real world, during a multiple document task, one prefers a large surface, like a dining table, to better visualise his/her documents. The spatial layout of documents could be considered essential for, among other processes, understanding their overall structure and relations among them. As with paper documents, digital documents inside an immersive 3D environment could be quickly brought to the user sight without overlapping with others. Moreover, the user would be able to place documents anywhere in the virtual space that fills his/her field of view and alternate his/her focus among them. This "enlarged" workspace would enable a better use of "external cognition", essential for managing such a complex task [17].

3 A 3D workbench for online digitized collections

In DL Web-based user's interfaces, browsing through collections and reading facsimiles are two separated tasks. This context break induces longer apprenticeship and navigation time within the interface. 3D interaction metaphors could be used to provide a continuous navigation space for these two tasks. The integration of the tools described in [10, 18] into one common workbench proved to be difficult, partly because of the inadequacy of the VRML language for such applications. Some design inconsistencies and limitations also appeared for 3D space navigation and for documents representation. In the collection interface of [10], the navigation is very constrained and each possible translations and rotations has a meaning for enlarging or focusing a research. Within the reading interface of [18], camera movements are free in order to allow arbitrary placement of windows. Zooming in and out (Z translations) is the most frequently used movement for this task, but other camera translations and rotations are allowed to permit the organization of various reading zones in space. The representation of books within 2D windows obviously did not capture the physical appearance of the documents. Showing book pages one after the other is problematical for texts that refer to plates, for example.

In this new version, the 3D scene is bounded in altitude by a virtual ground.

The camera position above the ground and its field of view are fixed. Slight camera changes on height are however possible, in order to obtain a different field depth and a satisfactory perspective. Books are represented and manipulated by a new metaphor shown in fig. 3 (left). It is made by a flat double page surrounded by a tripod which acts like a lectern. To ease interactions, the tripods are constrained to stand on the ground. Users can freely translate and rotate tripods on the workplane with a standard 2D pointer. To do so, the user selects the handle represented by the small green cylinder at the tripod's bottom. Closing/deleting a tripod can be done by a right click on the handle. More shortcuts for tripods placement should be defined. They could be accessed by a rotating menu around the handle.

Turning the pages interactively is also possible since a simplified 3D book structure, showing two pages at a time, is modeled. The pages can be turned either by clicking on the facsimiles or by using the vertical yellow axis, which acts as a standard scrollbar. The three usual navigation modes (go to next/previous, jump forward/backward, smooth fast forward/rewind) are also available with this scrollbar. Horizontal scrollbars are used to modify the zoom factor of overlying pages. When zoomed, images of the pages can be translated on the rectangular shape by a simple drag and drop. The thickness of the book should also be taken into account. Showing the books̀ edges in perspective would enable a faster understanding of the documentś structure. The edges could be augmented with user's bookmarks and highlights of chapters, as in [14].

Browsing tools such as bookshelves can be inserted in the 3D scene in different ways. A compromise must be found between the need for reduced navigation time (when switching from browsing to reading activity) and good screen estate management. Fig. 3 (left) depicts a "cockpit" architecture, similar in spirit to the Web Forager interface [9]. In this organization, the collection and the opened books share the same space. The collection is always visible and zooming on the bookshelves does not affect the tripods positions (i.e. opened books remain still). When a few tripods are being used together, the workbench is overloaded and large parts of the bookshelves are hidden. Two separated areas should therefore be defined. In the organization described in fig. 3 (right), the user can switch from one area to the other by a 180 degree rotation around the Yaw (Y) axis. The collection is accessible in the first zone and the user can move the point of view. In the second zone, the tripods are displayed on the ground.

4 The experimental setup

4.1 3D reading workbench

Each new generation of dedicated hardware for real-time 3D exhibits a considerable growth of processing speed, storage capabilities and, to less extent, screen resolution. Experiments described in [10] were in 1997-8 below acceptable frame refresh rates for fluid interaction (about 25 frames/s is the figure considered in the game industry). The on-board memory dedicated to textures management

was at that time too small for such applications. It is today quite possible to handle scenes that hold about 10K different books textures (which is the forecasted size of the CNUM collection). Pages are bigger image files, and at most a hundred can be stored directly in the graphic card's memory. This forbids "brute force" algorithms for fast page turning with large books. This problem has recently been studied in [14].

The first prototype described in [18] was mainly written in Java and relied on the 3D engine enclosed in a VRML browser. The amount of scripts required to implement the interaction behaviors of the 3D objects is important : over 90% of the demonstrator's code is made of Java scripts. Telling that VRML and Java is a useless couple to create interactive scenes is not the point. We do believe that almost every imaginable interaction can be coded this way. But it demands a long coding and debugging time. To ease the specification of such interactive scenes some extensions to VRML were proposed in [20] along with a description of the corresponding client navigator.

The prototype based on the book metaphor is a stand-alone application written in C++ and taking advantage of the OpenGL library [11]. All the 3D object were procedurally coded, as well as the texture mapping. The demonstrator includes a simplified collection shelf, tripods creation/deletion and page turning functions. This software has been demonstrated to IST professionals at JFT'2003 [21] and ECDL'2003 [22]. A wider audience has been reached during a two-weeks long demo session for the annual "Image par image" exhibition (Montreuil, France) in march 2003. In October 2004, a similar demonstration took place at the CNAM library for the "Science en fête" nation-wide event. Each time, a dedicated machine was installed for the 3D environment, while another offered access to the CNUM website. A time switch limited the session to a few minutes and user's actions were logged. We have concluded from these demonstrations that users reached quite quickly an understanding of the tripod operation. The page turning functions have been acclaimed by the general public and librarians (although not always by computer professionals). Observation of users drived us to modify the interface. The moving distance for tripods has been bounded : a tripod can not be bigger than the screen, nor too small when in the background. Collision detection has been improved to detect the users actions in the scene. This was done by ray casting from the 3D cursor. Scene lighting and objects shadows have been implemented to facilitate the positioning task of the tripods. If the solutions to improve the interface usefulness were evident, their implementation was heavy to do and time consuming (i.e. the ray casting had to be entirely hard coded).

4.2   Immersive display

We have chosen the VisionStation as the immersive display for our experience. The VisionStation is a hemispherical display that uses a projection-based hardware and software, for displaying images in a 160 by 160 degrees field of view (FOV). It consists of an hemispherical surface (1.5 m diameter) attached to a

Fig. 3. Left : the whole collection is rendered and accessible in the background. The tripod can be dragged and rotated along vertical axis by selecting the handle. Favourite books are shown in the foreground. Right : Organization of the workspace for reading and browsing tools.

small table. The user sits in front of this table where he/she may place the interaction input. A normal projector, coupled to a fish-eye lens, is installed into the table. Such a lens provides a 180-degree horizontal and vertical projection distributing equally the same amount of pixels all over the projection surface. Depending on the user's head position he/she may have almost all his/her field of view filled with the projection (fig. 4).

The VisionStation is not just a large screen. A specific API is necessary for adapting the 3D scene images to this visualisation device. Five simultaneous scene renderings are necessary in order to produce a FOV wider than 90 degrees (since traditional rendering distorts perspective when the FOV exceeds this angle). The renderings are blended onto a unique image ; this image is then spherically distorted so that, during the projection, it may appear correctly on the curved surface.

The main drawback of our setup is the low spatial resolution of the image displayed on the immersive screen. The spatial resolution is the resulting ratio between the projector's resolution and the size of the projection surface. In our case, due to the large projection surface and to the limited projector's resolution (1024 x 768 pixels), pixels become quite visible when user's head is at a correct position. The user may perceptually increase spatial resolution by moving away from the screen but this would also simultaneously reduce the user's FOV. Evaluations in [23] emphasize the importance of high spatial resolution for visualisation tasks where details play an essential role (such as scientific visualisations).

We have chosen a bifocal approach as a solution for this drawback. Bifocal visualisation offers two simultaneous views of the same scene. The second display, originally used only for viewing a single document per time, is a 19 inches LCD flat tactile screen with 1280 by 1024 pixels resolution. Thereafter it is a device

with a good spatial resolution that fits well our goals. It has been shown in [24] that using a higher spatial resolution device to show a small part of the image that is being showed by a lower spatial resolution device is an efficient method used in tasks not so different than ours. We have set up our software for rendering, simultaneously to the main image, a focus view corresponding to a small central part of the main view. This rendering is displayed in the small screen and works as a magnifying glass of our virtual environment. Images seen through this screen are notably clearer, crisper, and brighter than those displayed on the immersive display. The second screen lays on the VisionStation table so that the user's eyes may easily focus on it.

The prototype works satisfactorily well and it is easy to operate. Moving around the scene looking for a specific title has become a banal task. The VisionStation, unlike other immersive devices such as Head-Mounted Displays, is not exhausting for long-term use. Visualisation through the second screen offers greater comfort when more detailed information about the desired document is needed. Although the idea of virtually picking the documents is highly interesting, direct manipulation seems unfeasible because objects beyond the projection lens position cast shadows on the projected images.
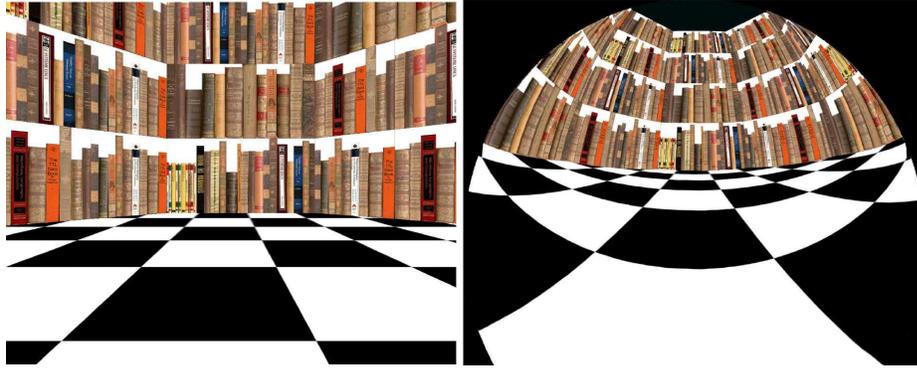
Fig. 4. 60 field of view (left) 160 field of view (right).

5 Future work

A lot of work was necessary for setting up the immersive visualization. Because the pitch is important on projection devices, images displayed were strongly aliased. Since the whole scene requires five renderings per frame, memory consuming processes such as oversampling were unfeasible. We have tested several alternative methods for reducing image aliasing. A visual comfort and slightly blurred images were reached using lower resolution textures for intermediate

Fig. 5. A working session with the immersive display and the focus screen.

renderings. MipMapping techniques also contributed to reduce image noise produced by dynamic textures transformations.

The software used in our prototype has already been demonstrated at several occasions and a usage evaluation is under development using the guidelines proposed in [25]. The physical workbench has not been tested since essential issues for this system remain undefined. How to switch documents from one display to another ? The documents will have different appearance and interaction properties depending on the device they are being displayed ? Should the scene rendered on the focus screen be spatially connected to the one on the context screen or should they display independent environments ? In addition, developing gesture-based interaction models are essential given the goals proposed for this research. Turning pages, zooming, panning, or even ripping off the book pages in order to create a different spatial layout are some examples of intuitive interactions. With such metaphorical interfaces, users would be able to directly manipulate the documents' facsimile.

References

1. Lesk M.: Practical digital libraries - Books, Bytes and Bucks Morgan-Kaufmann (1996)
2. Soergel D.: A framework for digital library research. Broadening the vision. D-Lib magazine 8(12) (2002)
3. Ramelli A.: Le diverse et artificiose machine (...). Parigi : in casa del'autore (1588) Also available online at CNAM http://cnum.cnam.fr/fSYN/fDY3.html
4. Fingerhut M.: Le site Web de la bibliothèque considéré comme un espace. Bulletin des bibliothèques de France. 45(3) (2000)

5. Christoffel M., Schmitt B.: Accessing libraries as easy as a game. in K. Brner, C. Chen (eds) Visual interfaces to digital libraries, LNCS 2539 (2002)

6. Plénacoste P., Lecolinet E., Pook S., Dumas C., Fekete J-D.: Zoomable and 3D Representations for Digital Libraries. Proc. of IHM-HCI 2001, Lille, France, 2001

7. Fox E., Kipp N. et al: Networked Digital Library of Theses and Dissertations : An International Effort Unlocking University Resources. D-lib magazine, (1997)

8. Rauber A., Bina H.: Visualizing electronic document repositories: drawing books and papers in a digital library. Proc. of IFIP VDB5, Fukuoka, Japan, May 2000

9. Card S. K., Robertson G., York W.: The WebBook and the Web Forager : An Information Workspace for the World-Wide-Web. Proc. of ACM CHI'96. Vancouver, Canada, April 1996

10. Cubaud P., Thiria C., Topol A.: Experimenting a 3D Interface for the Access to a Digital Library. Proc. of ACM DL'98. Pittsburgh, USA, June 1998

11. Cubaud P., Stokowski P., Topol A.: Mixing Browsing and Reading Activities in a 3D Digitalized Library. Proc. of ACM-IEEE JCDL'02, Portland, USA, June 2002

12. Carpenter L., Shaw S., Prescott A. (eds.): Towards the digital library : The british library initiatives for access programme. London, British lib., 1998

13. Chu Y.C., Bainbridge Jones D., Witten I.: Realistic books : a bizarre hommage to an obsolete medium ? Proc. of ACM-IEEE JCDL'04, Tucson, June 2004

14. Card S. K., Hong L., Mackinlay J. D., Chi E. H.: 3Book: A Scalable 3D Virtual Book. Proc. of ACM CHI'04. Vienna, Austria, April 2004

15. O'Hara K., Sellen A.: A comparison of reading paper and on-line documents. Proc. ACM CHI'97, Atlanta, USA, 1997

16. Mackinlay J. D., Heer J., Royer C.: Wideband Visual Interfaces: Sensemaking on Multiple Monitors. PARC Technical Report, 2003

17. Scaife M., Rogers Y.: External cognition: How do graphical representations work? Int. Journal of Human-Computer Studies, 45:185-213. 1996

18. Cubaud P., Topol A.: A VRML-based user interface for an online digitalized antiquarian collection. Proc of ACM SIGGRAPH Symposium Web3D'2001, Paderborn, Germany, Feb. 2001

19. Robertson G., Card S., Mackinlay J.: Information visualization using 3D interactive visualization. Comm. of the ACM, 36(4), pp. 56-71, April 1993

20. Topol A.: Interaction 3D pour les paysages informationnels. Conservatoire nat. des Arts et métiers PhD thesis, 2002

21. Cubaud P., Dupire J., Topol A.: Textes, images, volumes : les bibliothèques numériques au CNAM. Premières journées francophones de la toile (JFT). Tours, France, july 2003

22. Dupire J.: Digital libraries at CNAM : 1993-2003. Proc. of ECDL'03. Trondheim, Norway, Aug. 2003

23. Kasik D., Troy J., Amorosi S., Murray M., Swamy S.: Evaluating Graphics Displays for Complex 3D Models. IEEE Computer Graphics & Applications 22(3):56-64. 2002

24. Baudish P., Good N., Stewart P.: Focus plus context screens : combining display technology with visualization techniques. Proc. of UIST'01. Orlando, USA, Nov. 2001

25. Bowman D. A., Kruijff E., LaViola J. J., Poupyrev I.: 3D User Interfaces Theory and Practice. Addison Wesley, 2004

This article was processed using the LaTeX macro package with LLNCS style