SEMANTIC ZOOMABLE INTERFACE FOR MULTIMEDIA CONTENT

Daniele Panza

Università di Bergamo
Faculty of Engineering
viale Marconi 5
24044 Dalmine (BG) ITALY
daniele.panza@fastwebnet.it

Alexandro Sentinelli,
Andrea Vitali, Luca Celetto

STMicroelectronics, AST Labs
via Vignolina 17
20041 Agrate Brianza (MB) ITALY
{alexandro.sentinelli, andrea.vitali,
luca.celetto}@st.com

Gustavo Marfia,
Marco Roccetti

Università di Bologna
Dep. of Computer Science
via Mura Anteo Zamboni, 7
40127 Bologna (BO) ITALY
{marfia,roccetti}@cs.unibo.it

ABSTRACT

Without any doubt multimedia content has become essential in all aspects of our digital society, leading to the generation and storage of an unfathomable amount of digital media. In the Information Retrieval (IR) field the majority of research available to this date approaches all questions in terms of semantic tagging, indexing and feature extraction.

Although these are all fundamental steps in the design of any IR system, we believe that also an efficient human machine interface (HMI) can significantly improve the retrieval rate success at the end-user side. In our work we developed a Zoomable User Interface integrated with semantic algorithms dealing with media content, calling it a Semantic ZUI: we believe thath this approach can help browsing multimedia files in a seamless way, providing benefits for end users.

1. INTRODUCTION

Recent studies have investigated uses, features and drawbacks of common user interfaces for browsing, searching and retrieving multimedia content, especially for personal uses. A key point is designing interfaces able to use metadata (extracted from files or collected from other sources) to help users in their common task: organizing, browsing, searching, sharing. Graphical User Interfaces have evolved in their look, feel and graphic appearance, but their paradigm has roughly maintained the same approach over the years. We believe, instead, that the huge and constantly growing amount of media stored on our systems requires a fully scalable approch and this may involve exploring new kind of interfaces. Zoomable User Interfaces have scalability as one of their basis: the use of multi-representation objects allows to quickly move from a general overview to a detailed visualization of single objects. In academic environment, Ben Bederson et al. firstly explored the potentialities of ZUIs in several works [1, 2, 3]. In commercial applications, we have seen ZUIs rise to fame with Google Earth and Google Maps; Microsoft developed the Silverlight extension DeepZoom (formerly named Seadragon).

And ZUIs are making their way in many hand-held devices, like Apple's iPhone and Nintendo DS's web browser.

In this work we examine how the zoomable approach can be exploited together with a semantic algorithm and clustering techniques, and how the use of ad hoc layouts and dispositions can help browsing and retrieval: our efforts aimed at creating an advanced filesystem browser, starting from an existing prototype, as described in the next sections.

2. RELATED WORK

Ben Bederson firstly exploited zoomable interfaces in **Pad++** [1] and in **Photomesa** image browser [3]; Photomesa (and former **PhotoFinder** [4]) was used as base for experimenting several techniques of annotating, tagging and browsing images. Huynh et al [5] proposed a "time quilt" layout which combines advantages of timeline dispositions and space-filling treemaps. Hilliges et al [6] also used the zoomable approach: in their prototype of **PhotoHelix**, a touch screen table for photo sharing and storytelling, they layout clusters of images onto a spiral, allowing to zoom on them.

The starting point for our work is **Filemage**, a zoomable filesystem browser developed in [7]. FileMage is based on the Piccolo2D framework, which cover the basic functionalities needed for implementing ZUIs.

FileMage allows to browse the filesystem in a zoomable fashion: files and folders are disposed on a infinite zoomable plane, nested according to filesystem hierarchy. The user can navigate inside its files through the mouse: clicking on a file automatically adjust the camera's position and zoom on it; moving inside the hierarchy is intuitive and straightforward. FileMage can display text files, images (both scalar and vector) and video files. It manages to display a clean and intuitive view using multiple-representations objects: a small file is shown in a simplified way, and all its details are displayed only when zoomed. Likewise, a small folder does not show its content until it is enlarged.

FileMage shows how ZUIs can be effective for browsing large collections of hierarchical data (in this case, filesystem),
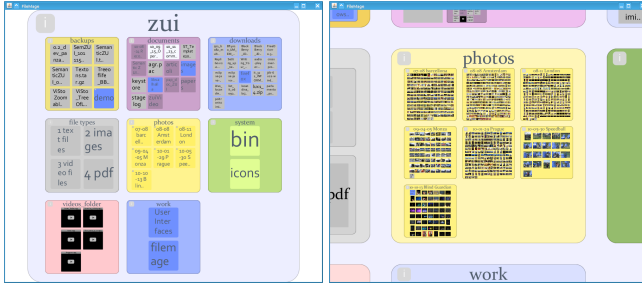
Fig. 1. FileMage browser interface showing a filesystem portion (left) and after a zoom-in on the "photos" folder (right)

offering a space-efficient visualization: with just a few clicks, users can jump from general overview to a detailed representation of a particular item. However, it lacks of search functionalities and tagging capabilities, and does not use particular layouts: files are disposed in a squared grid, ordered alphabetically, and all the objects in the same folder have the same size.

3. MOVING TO A SEMANTIC ZUI

FileMage has a very modular architecture, making it a perfect base for developing and testing semantic extensions: in the next sections we present the features we introduced.

3.1. Generating/Extracting Metadata

3.1.1. Manual and automatic tagging

FileMage was extended to allow the association of tags (keywords) to files. It was an obvious step for turning FileMage into a semantic interface: filesystems have a hierarchical structure, and tags are somewhat an orthogonal organization paradigm to hierarchy, besides being widely used (especially on the web). Tags may help overcome some implicit limitations of hierarchical organization, such as the imposition of a single order of categories, when many could apply (for example, this paper could go in docs/papers/semanticzui or semanticzui/docs/papers) or the growth in either depth or breadth, which slows down the browsing[8].

A user may tag a file simply typing a keyword for it or speaking it: we employed a **speech recognition** engine to recognize the user voice and extract a keyword from it. However, the tagging process cannot be limited to user-provided keywords: tagging is often perceived as a boring and time consuming operation, and users need to be encouraged to make a continuous use of tagging [9]. Some file types already include some metadata, like jpg's exif or mp3's tags, that can be extracted and used as tags. We employed the **Libvalhalla** library [10], a media scanner and parser library: it features "grabbers" able to retrieve files informations, from the files themselves and also from internet sources (automatic download of covers, lyrics and tags retrieval).

This kind of automatic tagging can already retrieve most useful informations for browsing a media collection, such as authors or albums names, partially relieving the user from an heavy task. In future development, we plan to expand the range of automatic tagging employing other algorithms, for example face and person recognition.

3.1.2. Multi-levels Visual Storyboards for videos

With the increasing proliferation of digital video contents, efficient techniques for analysis, indexing and retrieval of videos depending on their contents have become evermore important. YouTube is estimated to spend more than 11M$ per month for bandwidth transit expense [11] and according to Gill et al. [12] many YouTube sessions are interrupted when the content is not interesting to the end user, though the good video quality and the abundance of bandwidth. Instead of just one frame (usually the first one) and some words that most of the times are not sufficient/reliable to describe the content, the video storytelling (or storyboard) is a promising browsing tool to get a quick understanding/recalling of the semantic content of a video through visual examples.

We embedded a modified version of ViSto [13] a key frame extraction algorithm with the aim of enhancing the browsing performance of the ZUI in a Multimedia archive. ViSto is particularly interesting in our UI environment because it offers the user the possibility to choose the number of representative frames. As the ZUI is based logically on a hierarchical visualization it is possible to exploit the same approach by extracting video storyboard whose length is proportional to the zoom of the end-user.

3.2. Browsing

FileMage strictly reflects the filesystem hierarchical structure: we added three others organization structures, as explained in the next sections, and the possibility to filter the view. New structures arrange files according to their metadata: we developed a tag-based view and two time-based cluster organizations. In order to provide more flexibility, it is possible either to switch between organizations, or to mix them: the user can
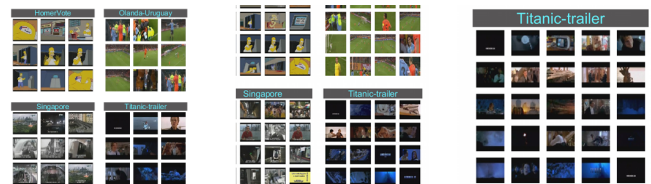

Fig. 2. From left to right, different stages of a zooming operation on a specific video storyboard, resulting in seeing more frames for the storyboard

browse in the filesystem hierarchy till a folder and organize its content by tags or by time. An organization may perform better with some types of content and not with others, or a user may prefer a method rather than others. Creating a system which adapts itself to the user needs and attitudes is the final objective.

3.2.1. Filtering by tag and tag-based view

Users can exploit tags in two ways: **filtering the view**, leaving only files and folders with a specified tag, or showing **files grouped by tags**.

Filtering can be performed in the same way as tagging: typing a keyword or speaking, using the speech recognition engine. After entering the keywords, nodes which do not possess it fade away.

Rearranging the view by tag displays each keyword as a folder, with related files inside it. As common web tag-clouds do, keywords associated to a large number of files are bigger than others. By tracking how many times a keyword is used to retrieve a file, we can determine its popularity: popular keywords (the ones used the most) are painted in brighter colors, whilst less common ones are darker.

3.2.2. Time-clustering organization

Another organizational method we implemented consists in arranging files in a **time-based hierarchy**, by exploiting a metadata common to every file: the "last modified" date. Files are shown in a dedicated layout which uses calendar-like widgets, grouping together files last edited in the same day. In this way is easy to retrieve a file if you remember when you used it, or to acknowledge the presence of old files not needed anymore.

This approach can be further refined if we consider a collection of personal photos: if we look at the distribution of time of still (saved in exif metadata by every digital camera), we notice that photos are not temporally equidistribuited, but usually come in bursts, of variable length. Most of the times, these groups correspond to specific events, such as a night
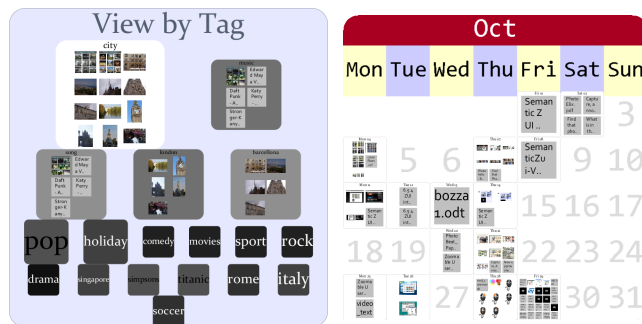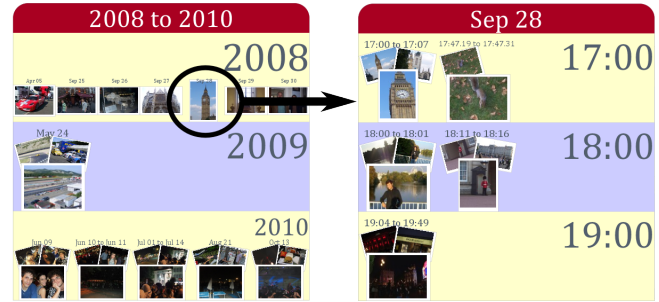

Fig. 4. Browsing "events" of photos grouped with a clustering algorithm based on EXIF metadata

with friends or a two-weeks holiday. Therefore, it is possible to profile a user collection and divide it in events: we believe that browsing "events" instead of photos can be effective.

A time-clustering algorithm has been developed as part of a thesis work[14] and later integrated into FileMage. The algorithm automatically identifies clusters of images belonging to the same event (that is, temporally close to each other), and it is flexible enough to work at any time scale, grouping photos that span across years or hours. For each cluster, a set of representative photos is chosen using the ViSto algorithm, adapted for a photo album. In the ZUI, at a small scale a cluster is represented using thumbnails of representative photos; when enlarged, its content is displayed.

4. THE VIDEO BROWSING DEMO

The semantic ZUI is a stand-alone software perfectly usable for browsing local collection of files. However, to demonstrate its flexibility, we integrated the ZUI in a wider application, a prototype for a video browsing tool based on standard web technologies which aims to be an inter- and intra-video browser. As said, the demo was developed for the P2P-Next project [15] as an interface for browsing a multimedia collection, especially videos; the initial version of the page was developed by BBC, which has strong interest in semantic features for video browsing purpose. We integrated the ZUI and packaged it as a Firefox plugin.

Navigation in the video collection is made through the ZUI, which is embedded as a Java applet in a container overlapped to the page. Videos are represented by their storyboards: when the user clicks on a storyboard's frame, ZUI container slides away and video playout starts from that frame. In the main area, beside the video and the standard progress bar, other tools are provided to navigate inside the video. A carousel of frames extracted with the ViSto algorithm[13] makes possible to jump at the most significant sections of the video. We are also experimenting "Semantic Timelines", special timelines highlighting the presence of a feature (such as, a face or a particular object) in specific sectors of the video.

Fig. 3. Files organized by tags (left) and by last modified date (right)

Fig. 5. video browsing demo (the ZUI is currently hidden)

There is also a button for recording a tag to be associated to the video currently seen; the same speech recognition engine described before is used.

5. CONCLUSIONS

In the last years user habits and default requirements are evolving dramatically. The market needs new kinds of UIs in terms of interactivity, intuitiveness, and semantic approaches. The aim of this work was to design a prototype of Semantic Zoomable User Interface (Sem - ZUI) plugged into a popular Internet browser. To the best of our knowledge, our Sem-ZUI represents the first contribution that integrates audio/visual semantic engines for tagging, organizing, extracting the metadata either already available or from the content itself, into a ZUI for file systems (a video is available for download at [16]). We hope that this work will contribute to the debate in the Information Retrieval community and encourage a look with a very broad perspective, ranging from semantic algorithms and metadata extraction to application-level content presentation design.

6. AKNOWLEDGEMENTS

7. REFERENCES

[1] Benjamin B. Bederson and James D. Hollan, "Pad++: a zooming graphical interface for exploring alternate interface physics," in *Proceedings of the 7th annual ACM symposium on User interface software and technology*, New York, NY, USA, 1994, UIST'94, pp. 17–26, ACM.

[2] B.B. Bederson and B. McAlister, "Jazz: An extensible 2d+ zooming graphics toolkit in java," *University of Maryland Computer Science Tech Report CS-TR*, vol. 4015, 1999.

[3] Benjamin B. Bederson, "Photomesa: a zoomable image browser using quantum treemaps and bubblemaps," 2001, UIST'01, pp. 71–80.

[4] Hyunmo Kang and Ben Shneiderman, "Visualization methods for personal photo collections: Browsing and searching," in *Proceedings of IEEE International Conference on Multimedia and Expo*, 2000, ICME2000.

[5] David F. Huynh, Steven M. Drucker, Patrick Baudisch, and Curtis Wong, "Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections," 2005, CHI'05.

[6] Otmar Hilliges, Dominikus Baur, and Andreas Butz, "Photohelix: Browsing, sorting and sharing digital photo collections," *Horizontal Interactive Human-Computer Systems, International Workshop on*, vol. 0, pp. 87–94, 2007.

[7] Silvio Moioli, "Libreria per interfacce utente zoomabili ed applicazione dimostrativa," First degree thesis, Università Degli Studi Di Bergamo, 2007.

[8] J. Hopkins, I. Vassileva, "Beyond keywords and hierarchies," *Journal of Digital Information Management*, vol. 3, 2005.

[9] Hyunmo Kang, Benjamin B. Bederson, and Bongwon Suh, "Capture, annotate, browse, find, share: Novel interfaces for personal photo management," 2009.

[10] "Geexbox valhalla," http://libvalhalla.geexbox.org/, accessed Dec 9, 2010.

[11] Dale C. Cheng, X. and J. Liu, "Statistics and social network of youtube videos," 2008, pp. 229–238, Enschede, The Netherlands.

[12] Zongpeng Li Anirban Mahantix Phillipa Gill, Martin Arlittz, "Youtube traffic characterization: A view from the edge," 2007, IMC'07.

[13] Marco Furini, Filippo Geraci, Manuela Montangero, and Marco Pellegrini, "Visto: visual storyboard for web video browsing," 2007, CIVR'07.

[14] Matteo Meroni, "Creazione di un visual storyboard per foto/video album," Università Degli Studi Di Milano, 2010.

[15] "P2P-Next project," www.p2p-next.org, accessed Dec 9, 2010.

[16] "Semantic zui demo video," http://db.tt/8BBunAR, accessed Dec 10, 2010.