REYKJAVÍK UNIVERSITY
HÁSKÓLINN Í REYKJAVÍK

Breaking out of the Shoebox: Towards Having Fun with Digital Images

Kári Harðarson, Björn Þór Jónsson, Laurent Amsaleg

RUTR-CS07002 — August 2007

Reykjavík University - School of Computer Science

Technical Report

ISSN 1670-5777

REYKJAVÍK UNIVERSITY
HÁSKÓLINN Í REYKJAVÍK

Breaking out of the Shoebox: Towards Having Fun with Digital Images

Kári Harðarson*, Björn Þór Jónsson*, Laurent Amsaleg†

Technical Report RUTR-CS07002, August 2007

Abstract: The tremendous growth of personal digital image collections calls for new ways to access these images in a useful and interesting manner. We propose that such a browsing interface should combine features such as multi-dimensional pivoting and space travel-like exploration. This report describes an image browsing prototype and our experience from using it with two different real-life data sets. Our conclusion is that while the prototype has shortcomings, this is a very promising research direction that merits further exploration.

Keywords: Digital images; Browsing; Searching; OLAP; Exploration.

(Útdráttur: næsta síða)

This work was partially supported by Icelandic Research Fund grant 070005021, INRIA Eff[2] Associate Teams grant, and an ÉGIDE Jules Verne travel grant.

* Reykjavík University, Kringlan 1, IS-103 Reykjavík, Iceland. karih/bjorn@ru.is
† IRISA–CNRS, Campus de Beaulieu, 35042 Rennes, France. laurent.amsaleg@irisa.fr

Brotist úr viðjum skókassans: Hvernig má skemmta sér með stafræn myndasöfn

Kári Harðarson, Björn Þór Jónsson, Laurent Amsaleg

Tækniskýrsla RUTR-CS07002, Ágúst 2007

Útdráttur: Ljósmyndasöfn einkaaðila fara sístækkandi. Þetta kallar á nýjar aðferðir við að skoða ljósmyndir á gagnlegan og ánægjulegan hátt. Við teljum að slíkt skoðunarviðmót ætti að nýta aðferðir eins og margvíða skoðun þar sem notandinn getur sjálfur valið víddir,og skoðun á "geimi" eins og gert er í tölvuleikjum en þar sem geimurinn er fullur af myndum. Þessi skýrsla lýsir frumgerð að myndaskoðara og reynslu okkar af notkun frumgerðarinnar við tvö raunveruleg myndasöfn. Niðurstaða okkar er að þótt frumgerðin hafi galla sé þetta álitleg rannsóknarstefna sem kalli á frekari skoðun.

Lykilorð: Stafrænar myndir; Skoðun; Leit; OLAP; Könnun.

(Abstract: previous page)

Contents

1 Introduction

In recent years, the world has seen a tremendous increase in the capability to create, share and store digital images. As a result, personal image collections are growing at an astounding rate and it is clear that in the future individuals will need to access tens of thousands, or even hundreds of thousands, of digital images. It is therefore imperative to start studying ways to access these images in a useful and interesting manner. What is needed is software that will allow users to seamlessly organize, search and browse their images, while having fun at the same time.

Many products and prototypes exist that attempt this integration, such as ACDSee (www.acdsystems.com), Picasa (www.picasa.com), PhotoMesa [Bed01] and QBIC [FBF$^+$94]. Unfortunately, much of the functionality of these systems amounts to a digital "shoe-box", where users can browse images based on simple attributes, such as time, disk location or user-supplied keyword classification, and possibly search by these same attributes. Presentation is typically linear and the contents of the images are not used to guide the search and presentation. Notable exceptions are the El Niño [SJ99] and PIBE browsers [BCP04], which use the contents of images to determine the location on screen and to group together similar images into clusters, and allow users to modify the similarity relationships. A study by Rodden and Wood has shown that users are not happy with the functionality provided by photo browsers [RW03]. All these interfaces, even El Niño and PIBE, are lacking in freedom of movement through the image collection, making the browsing experience a tedious one, much like shoebox browsing of paper photos.

Each image may be described by a number of attributes, based on image contents and image meta-data.[1] Some of these attributes may be linear or spatial, such as time and location of taking the image, while others may be textual, hierarchical or categorical. Similarity ranking of search results may even be considered as an attribute of images. These attributes may be considered dimensions in an image hyper-space, which we must be able to traverse dynamically to fully enjoy our digital images. In on-line analytical processing, multi-dimensional data is dealt with by considering a few dimensions at a time and pivoting between dimensions when necessary. In advanced computer games, large three-dimensional worlds are explored by simulating space-travel. Both approaches have been very successful in keeping their users occupied and focused on their task for a long time. We propose that a browsing interface for images should merge these features into a multi-dimensional interface that allows flexible space-travel like exploration of the image hyperspace.

In order to begin exploring the possibilities of such a browsing interface we have implemented a prototype, using the Partiview browser [SL04], which allows us to browse images in a three-dimensional space. The dimensions may be based on image contents and image meta-data and different dimensions may be combined in an arbitrary manner. In this

[1] Some image meta-data, such as camera and time information, is stored in so-called EXIF headers (EXtended Image Format); see http://www.jeita.or.jp/english/.

report we briefly describe this prototype and our experience with its application to two different real-life data sets. Our conclusion is that while the prototype itself has shortcomings and should be abandoned, this is a very promising research direction that merits further exploration.

2 Related Work

Existing image browsers can be divided roughly into commercial browsers and research prototypes. In this section we describe examples from each category and end with a discussion of common features and flaws.

2.1 Commercial Browsers

There is a plethora of commercial (licenced and shareware) browsers available, most of which provide the user with a similar set of features. In our opinion, the best such browser is currently PicaJet (www.picajet.com). PicaJet divides the screen into three columns. The middle column displays a 'light table' with thumbnails of selected images. The left and right hand columns, on the other hand, are used to select subsets of the images for browsing.

The left column displays a hierarchy of categories. 'People', 'Places' and 'Events' are suggested as root categories, but the user can add categores at will. The right column typically displays a time line, which the user can zoom in on to see resolution of years, months, days or even hours. The number of images at each granularity is summarized on the time line. The right column can also be used to select by ratings and file locations. Both columns essentially provide a filtering mechanism, which can be used to show only the images belonging to a particular subset of all images.

Needless to say, the browser has many other nice features, such as meta-data and photo editing and slide-show generation, which are required in modern commercial browsers. The browsing interface, however, does not allow dynamic choice of dimensions to browse by and fails to take proper advantage of the two-dimensional nature of the screen.

2.2 Research Prototypes

Several research prototypes have been proposed. These browsers are typically geared towards demonstrating the viability of a particular research topic, but do not attempt to provide a complete solution. For example, PhotoMesa [Bed01] uses Quantum Treemaps [BSW02] to maximise screen utilization and provides a zoomable user interface for navigation of the folder hierarchy. No attempt is made to consider other browsing dimensions, such as categories or time.

Ramesh and Jain were the first to point out the dichotomy between browsing and searching and proposed to create an interface to integrate the two [SJ99]. The resulting El Niño browser provides a direct manipulation interface where the user operates on the database's similarity measure by placing images in a position that reflects their mutual similarity in the current context. Rather than truly integrating searching and browsing, however, this is essentially a query refinement interface where the ranking function can be modified.

While such a query refinement interface requires that the database engine be able to adapt its similarity measure, El Niño did not provide an adaptive data structure. This problem was addressed by the PIBE Browser [BCP04]. PIBE initially creates a hierarchical browsing structure by clustering images by their similarity, choosing representative images for each cluster. Users can then alter the similarity measure by re-arranging images on the screen, resulting in a modification of the browsing structure. The browsing structure is persistent, allowing modifications by users to survive a restart of the system. Again, however, the PIBE browsing interface does not allow dynamic choice of dimensions to browse by and fails to take proper advantage of the two-dimensional nature of the screen.

2.3 Summary

In summary, the best-of-breed features of modern photo browsers include: efficent thumbnail presentation; packing thumbnails onto the screen for good overview; choosing representative images for clusters of images; and browsing via a zoomable user interface. All current photo browsers, share several design flaws:

- They only use limited and static dimensions for browsing (typically restricted to time, folder location, and categories). While image content and image metadata may on occasion give the user interesting insights, such dimensions are generally not used.
- They only allow using a single dimension at a time, for example the timeline. Thus the two-dimensional capability of the computer monitor (or three-dimensional, with clever graphics algorithms) is not used as well as in OLAP systems.
- They only allow for relatively limited, text-based annotations. Support is needed for quantitative annotations, in our opinion.
- They have limited search capabilities and completely separate the search and browsing functions. When, for example, annotations are used to search the image collection, a sub-set is displayed which is considered to match the search. Previous browsing information is lost; in effect the browser stops being a browser.

Our long term goal is to address all these flaws.

3 The Image Browser Prototype

3.1 The Partiview Browser

Our image browsing prototype is based on the Partiview browser [SL04], which was originally designed to browse images of galaxies. We chose to use the Partiview browser, because: 1) using Partiview, three-dimensional data sets can be displayed and traversed; 2) it has

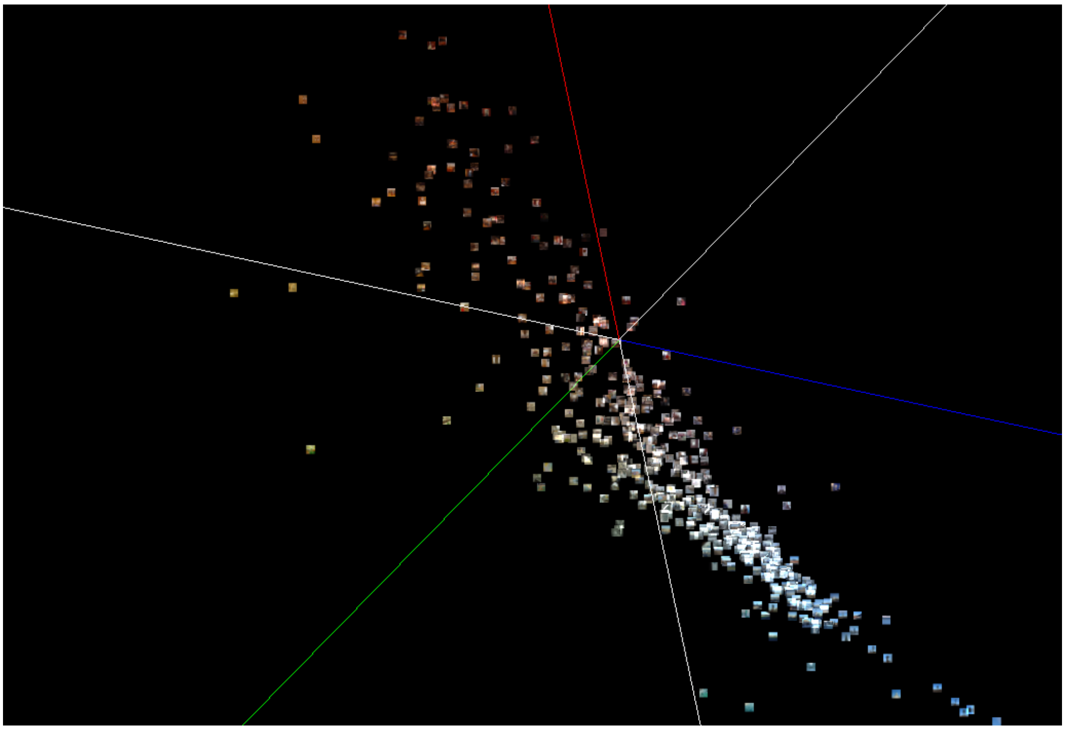

Figure 1: Overview of family album collection.

many features for controlling the display of images, such as resizing; and 3) it is tested, open-source software[2] that could be used to quickly prototype the navigation. Unfortunately, it is missing several features that we believe to be essential, such as pivoting between dimensions and viewing image originals, but as a first prototype the benefits outnumber the disadvantages.

The Partiview browser offers a three-dimensional view of a data set of image thumbnails. Once the data set is loaded, the user can navigate through the three-dimensional space using a combination of mouse and keyboard commands. The space can be rotated, but the dimensions cannot be stretched. A fourth (categorical) dimension may be assigned as a filter, and the Partiview interface allows the users to select some or all of the categories to appear on screen. Partiview does not allow the user to view the original images, but the images may be labelled with the file name, which facilitates detailed inspection of images when needed.

[2] Available at http://niri.ncsa.uiuc.edu/partiview/

3.2 Using Partiview

In order to prepare our data sets for Partiview, we use a set of python scripts and the ImageMagick library. Our two data sets contain JPEG images taken from a digital camera and web-camera. These images are converted into thumbnails of size 128x128 in the SGI RGB-format used by Partiview. At the same time, several content-based descriptors are computed from the image contents, and the EXIF headers of the images are used to yield image meta-data descriptors. The descriptors used are described below.

All these descriptors, along with the file name, are combined into a comma-separated list and stored along with the thumbnails. While this process takes some time, depending on the number of images, it is a one-time operation.

Once the descriptors and thumbnails are ready, another python script is used to assist the user in selecting the dimensions to be used and extract the dimension information into a control file, which is read by Partiview. Once this operation has been performed, the Partiview browser can be started. For our larger data set of almost nine thousand images, Partiview takes more than five minutes to start on our lap-top computer, as it must read the data set into memory.

3.3 Dimensions of the Image Hyperspace

We have defined over thirty different dimensions, based on the image meta-data and contents. Among the meta-data based dimensions are the f-stop value, shutter speed, film speed, the absolute brightness (f-stop $\times$ shutter speed / film speed), exposure time, flash used (binary), focal length, image height and width, and the date (and many derived metrics such as day, month, and year). Among the content-based descriptors are the weighted average color values for the red, green and blue colors, the ratio of these color values, as well as numbers which indicate which red, green and blue color levels are the most common in the images.

4 The Browsing Experience

We have experimented with two very different data sets. The first data set is a private collection of 474 family images, which are taken in winter. The second data set is a collection of 8,661 images taken from a web-camera in Reykjavík over the span of a whole year. The data sets are described in more detail below, along with the authors' experiences with browsing the data sets.

4.1 Collection 1: The Family Album

The first image collection we studied, the FAMILY data set, is a collection of 474 family images, taken in various places during winter in Iceland. For the presentation we chose the ratio of red, green and blue colors in the images. Figure 1 shows an initial view when approaching the 'cloud' of images from space. We can see that most images are located near the middle, having roughly equal amounts of red, green, and blue colors. There are,

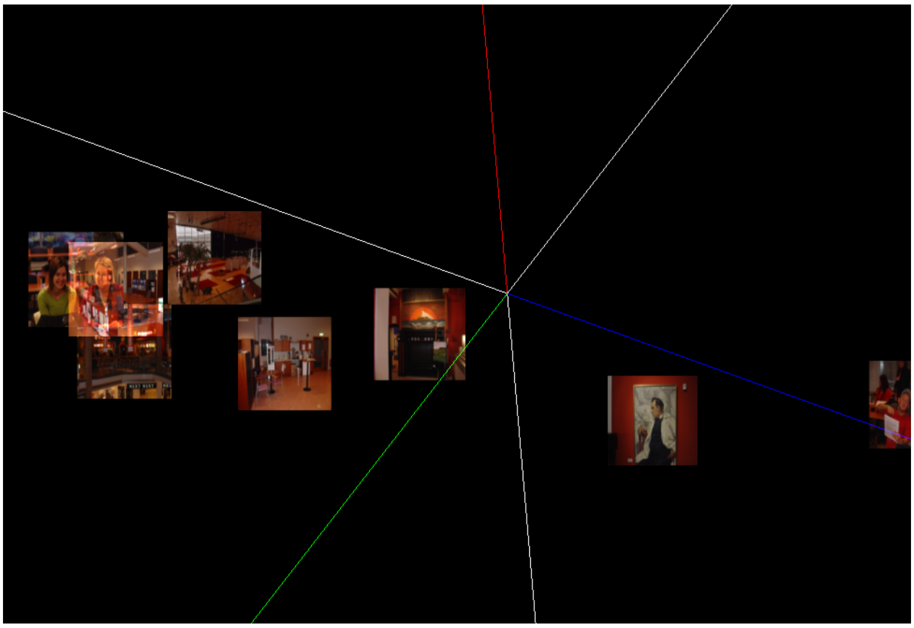

Figure 2: A close-up of the 'red wing' of the family album collection.

however, some outliers that form red and blue 'wings' in the data set, while only a few images have a majority of green colors (recall that the images are taken in winter).

Further inspection of the 'red wing' reveals the pictures seen in Figure 2, which are mostly taken indoors in poor lighting conditions. Similar investigation of the blue wing, seen in Figure 3, reveals a great deal of blue sky, and so on.

Looking back at Figure 1, however, we note the many images clustered together in the middle. Clearly other axes are necessary to better separate and understand these images; being able to pivot between dimensions would be handy. Being able to specify filtering ranges on dimensions that are not being displayed ('dicing' in OLAP terminology) would also help to hide images that are not of interest at the time.

4.2 Collection 2: The Web-Camera Images

The second image collection we studied, the web-image data set, is a set of 8,661 snap-shots from a web-camera located in Reykjavík. One snapshot was taken every hour for a whole year.[3]

[3] The number of images should be $24 \times 365 = 8760$, but occasionally the web-camera was down or the images were corrupted for some reason, and hence 99 images are missing.

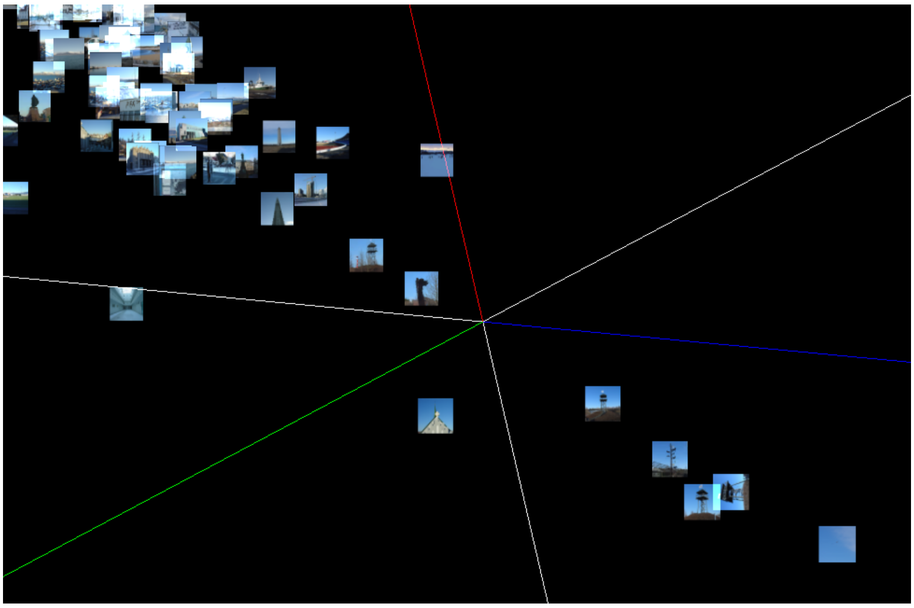

Figure 3: A close-up of the 'blue wing' of the family album collection.

Since this data set has uniform location and subjects, we focus on time and daylight aspects. Because the images are snapshots from a web-camera, the EXIF headers are missing, so the capture dates are used to generate the time dimensions; we have chosen the hour within the day and day within the year to visualize the time element of the data set. Since we do not have brightness information, we used the weighted sum of the blue color to indicate the daylight in the pictures; nighttime pictures have low levels and daytime pictures high levels.

Since Iceland is very close to the arctic circle, the difference between daylight in the summer and winter is very pronounced, with long days in summer and long nights in winter. Additionally, Iceland uses GMT, although it is located about 90 minutes west of Greenwich. This means that 'noon' occurs at about 13:30 in the afternoon and 'midnight' at about 1:30 at night.

Figure 4 shows an overview of the web-image collection, where the day of the year is viewed horizontally in the picture and the time of day vertically, while the brightness is directed 'out of' the picture. The figure clearly shows the sine-curve of the daylight, with the summer nights being quite bright and winter nights very long. The figure also shows how the brightness clearly dips at dusk all year round. Finally, with careful study the figure furthermore shows how noon occurs later in the day and midnight later in the night.

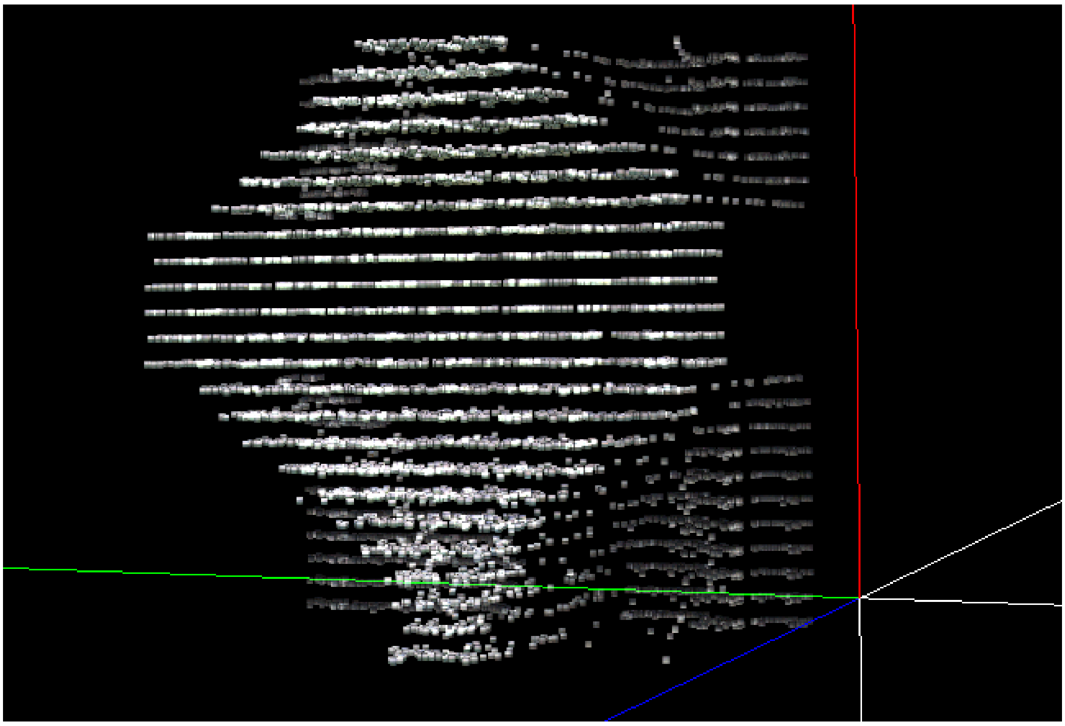

Figure 4: Overview of the REYKJAVÍK data set. The horizontal axis shows the time within the year, while the vertical axis shows the time within the day. The figure clearly shows the variance in daylight.

5 Discussion

The first observation we can make, is that we have had a great deal of fun looking at our data sets in this manner. Being able to fly through the data or view it from a perspective, as well as being able to look at different dimensions, has generated many new insights into the pictures.

Additionally, using the prototype with its limited functionality, has pointed out many issues that we would like to address in future work.

- There is no support for dimensions that are non-linear, such as folder location which is hierarchical in nature. Additionally, work must be done to add support for nearest neighbor ranking based on semantics, such as facial similarity or similarity based on local descriptors or objects in the images. Linking image content with other information, such as height or age of people in the images, might yield interesting graphics. Combining all these dimensions, especially with pivoting, will be a challenging task.

- Pivoting of dimensions is not possible. Since there may be many useful dimensions available, being able to dynamically switch dimensions is very important. Additionally, being able to specify filters on dimensions, regardless of whether they are currently used in the display, is a necessary feature.

- The interface should allow axes to be stretched dynamically. Unlike regular space, dimensions have different ranges and there may be dense clusters where stretching the axes would be useful to separate images. On a similar note, images that are very similar may disappear behind one another, which should somehow be avoided. One method might be to rotate between the images or to tile them together.

- The image hyperspace is a big place where it is easy to lose sight of the images. Unlike many computer games, the user can only see straight forward, which means that finding the images again may be very difficult. While it is possible to manually enter a command to move to the center-point of space, it would be preferable to prevent the user from losing sight of the images in the first place.

- The user should be able to view the original image by clicking on the thumbnail in the browser.

- The performance of Partiview is not suitable for the application, as all images are loaded into memory before the user can start browsing. Loading the images dynamically, as they enter the horizon, would be preferable. Facilitating dynamic loading of large data sets, in combination with pivoting and other features, will be a difficult task.

- So far, we have worked with a sizable number of images. We have found that it is easy to get lost in a sea of images. Since we expect that in the future much larger collections must be browsed, support for aggregating images along the horizon is required. For distant images, this could be done by simple counts, but for nearby images, this aggregation must be performed based on image contents. How to aggregate images dynamically in the multi-dimensional space, while efficiently supporting arbitrary pivoting, is another difficult question.

- Additionally, we have found that the EXIF standard is being interpreted differently by different manufacturers. Software from one manufacturer may not be able to open images taken with a camera from another manufacturer. Some fields in the EXIF standard are not being used, such as the distance to subject. It will be important to handle such practical issues.

- Given that GPS chips have become very cheap, GPS co-ordinates, direction and distance to subject will become key fields in determining the location of the subject of the image. Having this information will allow map-like presentation of vacations; with time as the third dimension, it will be interesting to trace routes.

The conclusion that we draw from this work is that the approach of combining pivoting, space-travel, and image aggregation has great potential for impact, but much more work is needed before a viable interface can be presented. Such work, of course, must draw on human-computer interaction theories and practice.

6 Conclusions

This report has described an image browsing prototype, based on the idea of exploring the image hyperspace via virtual space travel. We have presented an initial prototype and presented our experience with two different data sets. While we have listed many research problems that must be solved to make the approach viable in practice, we strongly believe that this approach holds great promise for allowing users to have fun with their images.

References

[BCP04] I. Bartolini, P. Ciaccia, and M. Patella. The PIBE personalizable image browsing engine. In *Proceedings of the CVDB Workshop*, pages 43–50, Paris, France, 2004.

[Bed01] B. B. Bederson. PhotoMesa: A zoomable image browser using quantum treemaps and bubblemaps. In *ACM Conference on User Interface and Software Technology*, pages 71–80, Orlando, FL, USA, 2001.

[BSW02] Benjamin B. Bederson, Ben Shneiderman, and Martin Wattenberg. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Transactions on Graphics*, 21(4):833–854, 2002.

[FBF+94] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, 1994.

[RW03] K. Rodden and K. R. Wood. How do people manage their digital photographs? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 409–416, Ft. Lauderdale, Florida, USA, 2003.

[SJ99] Simone Santini and Ramesh Jain. The "El Niño" image database system. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, pages 524–529, Florence, Italy, 1999.

[SL04] D. Surendran and S. Levy. Visualizing high dimensional datasets using Partiview. In *IEEE Symposium on Information Visualization (INFOVIS)*, page 20, Austin, TX, USA, 2004.

REYKJAVÍK UNIVERSITY
HÁSKÓLINN Í REYKJAVÍK

School of Computer Science
Reykjavík University
Kringlan 1, IS-103 Reykjavík, Iceland
Tel: +354 599 6200
Fax: +354 599 6301
http://www.ru.is