

Capture, Annotate, Browse, Find, Share: Novel Interfaces for Personal Photo Management

Hyunmo Kang

Human-Computer Interaction Laboratory,
University of Maryland Institute for Advanced Computer Studies, College Park, MD

Benjamin B. Bederson

Human-Computer Interaction Laboratory,
Computer Science Department, University of Maryland
Institute for Advanced Computer Studies, College Park, MD

Bongwon Suh

Palo Alto Research Center, Palo Alto, CA

The vision of ubiquitous digital photos has arrived. Yet, despite their broad popularity, significant shortfalls remain in the tools used to manage them. We believe that with a bit more creativity and effort, the photo industry can solve many of these problems, offering tools which better support accurate, rapid, and safe shared annotations with comfortable and efficient browsing and search. In this article, we review a number of projects of ours and others on interfaces for photo management. We describe the problems that we see in existing tools and our vision for improving them.

1. INTRODUCTION

The days when people debated the relative merits of film versus digital imagery now seem almost quaint. And with hindsight, film seems destined to have been a chapter in history along with LP vinyl records—a temporary physical analog recording medium. Although some aficionados may still prefer some qualities of those dust-gathering mediums, the advantages of digital media have become clear. The rapid and inexpensive ability to edit, annotate, search, share, and access has brought digital media to ubiquity.

And yet, with all this promise, shortfalls remain in the overall user experience. How many of us have replaced old unlooked-at shoeboxes of prints with unlooked-at digital archives of image files? How much has our ability to find a particular set of

Color figures are available at <http://www.cs.umd.edu/hcil/ben60>

Correspondence should be addressed to Hyunmo Kang, AVW 3211, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.
E-mail: kang@cs.umd.edu

photos really improved (i.e., can you find those photos of a visiting uncle with your sister when they were children?) How do we record stories from our parents describing family photos? How do we make sure those stories stay with the photo in question and get distributed to all copies of that photo within the family? Perhaps of most concern, how do we ensure that these annotations stand the test of time and remain accessible as computers, file formats, recording mediums, and software change?

These changes are all happening within the context of human behavior, which does not change so rapidly. People still like immediate gratification: take pictures rapidly, print them, and share in social settings. Some people spend a lot of effort creating photo albums or “scrapbooking.” Of course, many do not. Understanding which behaviors are fundamental and which are side effects of current technology is crucial, because this understanding can and should influence where researchers spend their effort.

We explore these issues and more with much of the intellectual motivation coming from Ben Shneiderman, our close colleague who has pushed for a deeper understanding of and better support of photo management for more than 10 years. His personal photo archives document the field of human–computer interaction (HCI) going back to its beginning. He regularly shares those photos with great enthusiasm to visitors, motivating and exciting all of us—largely because of the care and consistency he has applied to annotating and organizing his photos. He regularly pulls up old photos of lab visitors showing everyone what they worked on 5 or 15 years ago (and of course, showing what they looked like too!). His early exploration of tools to support photo management (with co-author Kang) led to PhotoFinder (Kang & Shneiderman, 2000), and the ensuing PhotoMesa tools (Bederson, 2001; Bederson, Shneiderman, & Wattenberg, 2002). His personal interest helped inspire the authors of this article as well as other lab members to pursue the development of approaches and software to improve all of our user experiences when managing photos.

This, of course, all happened during a time of tremendous commercial activity in this area. There are wildly popular photo sharing Web sites, such as Flickr (<http://www.flickr.com>; Yahoo!), Picasa Web Albums (<http://picasaweb.google.com>; Google), Snapfish (<http://www.snapfish.com>; HP), Shutterfly (<http://www.shutterfly.com>; Shutterfly), and PhotoBucket (<http://www.photobucket.com>; Photobucket), as well as equally well-used desktop photo applications such as Google Picasa (<http://www.picasa.com>), Adobe PhotoShop Album (<http://www.adobe.com/products/photoshopalbum/starter.html>), and ACDSee (<http://www.acdsee.com>). The two approaches (desktop application and Web site) are interesting to look at because they each offer distinct advantages to users. For example, Web sites are available anywhere and facilitate sharing, whereas desktop applications are faster, support higher resolution photos more easily, provide local ownership of photos, and offer richer interaction capabilities. It is interesting that each approach is gaining characteristics of the other. Web applications begin to offer dynamic, interactive content rather than static html pages through AJAX and Flash technologies. In addition, they often include plug-ins to ease uploading or improve performance, and some offer APIs to enable desktop applications to access their data directly. At the same time, many desktop applications are offering Web capabilities such as sharing.

Yet even with this commercial activity, the full potential of personal photo management has not been reached. There is the opportunity for richer annotation interfaces, automated content analysis, improved sharing, and more creative organizational strategies. Our hope is that more photos end up with better metadata, enabling faster, easier, and more accurate and enjoyable retrieval and use.

In this article, we look at some of the key activities and behavior patterns of personal photo users and examine how innovative user interfaces have the potential to enhance users' power, satisfaction, and control in managing and exploring their images. Starting with a close look at annotation, we examine how a combination of manual and automated techniques can improve how people associate metadata with photos. We then look at how the resulting richer metadata can enable better interfaces for searching and browsing photos. Finally, we end with a discussion about the importance of sharing photos and how new interfaces enable that.

2. GUIs FOR ANNOTATION

An essential question is, How valuable is photo metadata? Our own assessment of user needs (Shneiderman & Kang, 2000) coupled with reports from other researchers (Chalfen, 1987; Naaman, 2005; Rodden & Wood, 2003), and our personal experience come together on this. They indicate that the photo metadata such as dates, locations, and content of the photos (especially people's names) play a crucial role in management and retrieval of personal photo collections.

However, in spite of the high utility of the photo metadata, the usability of software for acquiring and managing the metadata has been poor. The manual photo annotation techniques typically supported by photo software are time-consuming, tedious, and error prone, and users typically put little effort into annotating their photos. In fact, the industry attitude tends to be that because users do not annotate their photos very much, it is not necessary to spend much energy adding good support for it. However the success of keyword labeling, so-called *tagging*, systems such as del.icio.us (<http://www.del.icio.us>) and Flickr hints that users indeed want to make annotations when reasonable utility and usability are supported. We believe that the photo annotation has enough utility for some users and it is the usability of software that needs to be improved.

In this section, a few innovative approaches are presented to show how interaction and graphical user interface (GUI) design can improve the usability of the photo annotation process when they are based on careful understanding of users' behavior and usage pattern. In addition, we explain how those designs have been evolving over time to support a broader range of annotation tasks by combining the accessible technologies with the analysis of users and their needs.

2.1. Advanced Manual Annotation: Direct Annotation

Because annotations based on automated image content analysis are still limited, we developed an advanced manual annotation mechanism that can significantly reduce users' annotation workload under certain circumstances. From the

observations of personal photo annotations (Shneiderman & Kang, 2000), we found that there were three interesting characteristics that could be useful for our interaction design.

- Personal photo libraries often contain many images of the same people at different events. In the libraries we looked at, we typically found 100 to 200 identifiable people in several thousand photos. Furthermore, the library has a highly skewed distribution with immediate family members and close friends appearing very frequently.
- Textual search often doesn't work reliably because of inconsistency in names with misspellings or variants (e.g., Bill, Billy, William).
- Lists of names of people appearing in photos are often difficult to associate with individuals, especially in group shots. Textual captions often indicate left-to-right ordering in front and back rows, or give even more specific identification of who is where. However this approach is tedious and error prone.

Based on these observations, we collaborated with Ben Shneiderman to develop the concept of *direct annotation* (U.S. Patent #7010751), which is a technique using selectable, draggable labels that can be placed directly on the photo (Shneiderman & Kang, 2000). Similar interfaces have also appeared recently on Flickr and MySpace. Users can select from a scrolling or pop-up list and drag by mouse or touch screen. This applies direct manipulation principles (Shneiderman, 1983, 2005) that avoid the use of a keyboard, except to enter a name the first time it appears. The name labels can be moved or hidden, and their presence is recorded in the database or in the header of an image file in a resolution-independent manner. The relative location of the target is stored based on an origin in the upper left-hand corner of the photo with the point in the range (0, 0) – (1.0, 1.0) corresponding to the full image. This approach not only associates a name with a position on the photo but also ensures that each name is always spelled the same way.

This simple rapid process also enables users to annotate at any time. They can add annotations when they first see their photos on the screen, when they review them and make selections, or when they are showing them to others. This design, which supports continuous annotation encouraged users to do more annotation, especially in collaborative situations such as with PhotoFinder Kiosk (Kules, Kang, Plaisant, Rose, & Shneiderman, 2004; Shneiderman et al., 2002). In a public setting, visitors of PhotoFinder Kiosk added 1,335 name labels using direct annotation while adding 399 captions using traditional type-in method.

The direct annotation mechanism was revised later so that it enables users to define their own categories such as activities, events, locations, objects in a photo in a hierarchical way (Figure 1) in addition to person names. A few alternative and complementary direct annotation mechanisms such as split menu annotation, hotkey annotation, and pop-up menu annotation have also been designed and developed to accelerate the annotation process. These are described in more detail in the following subsection.

A pilot experiment was conducted to see if the direct annotation method improved the annotation process in terms of annotation time and users' subjective preference compared with the traditional caption method or the click-and-type



FIGURE 1 The revised direct annotation mechanisms as implemented in PhotoMesa, the successor to PhotoFinder. *Note.* Users can add a caption under a photo, or add a particular attribute such as favorite (a yellow star on the bottom left of photo) or hidden. In addition, labels can be dragged from the list of people (on the top left) or from the user-defined category tree (on the bottom left). A label can be directly placed on the photo to represent where the individuals or objects are located within the photo.

method (Goto, Jung, Ma, & McCaslin, 2000; Jung, 2000). Forty-eight volunteers participated in a within-subject design, whereby each participant attempted an annotation task (20 names in five photos) on each system. The direct annotation method was significantly preferred, but no significant difference was found in the mean annotation time. A more extensive user study might help identify under what circumstances (e.g., number of total labels, average number of people in a picture, pattern of people's appearance in a personal photo collection, etc.) different annotation mechanisms work better than others in terms of completion time, error rate, users' satisfaction, or confidence.

2.2. Enhanced Direct Annotation: Bulk Annotation

The direct annotation mechanism was enhanced through a series of design cycles to support more efficient and productive annotation. Perhaps the most notable

one was applying the *bulk annotation* technique, which lets users annotate several photos with a single interaction (Kuchinsky et al., 1999). Bulk annotation is especially useful when users have a set of photos with the same person or group of people involved in the same events. We thus designed and developed several variations of direct annotation and bulk annotation mechanisms to accelerate the annotation process.

The split menu annotation mechanism (Figure 2a) was designed to minimize users' time of scrolling the list to find correct labels to be used for photo annotation. The split menu featured a resizable splitter bar so that the number of the most frequent labels displayed in the top window (Sears & Shneiderman, 1994). The scrollbar was removed from the top window, whereas the bottom half retained its scrollbar. The split menu raises interesting questions about what kind of automatic algorithms should be used to predict users' future access and facilitate rapid annotation.

Because some annotations get used so much more frequently than others, we designed the interface so that each label (either person's name or categories) can be associated with a hotkey (Figure 2b). After the key is assigned to a particular label, users can press that key whenever the mouse is over a photo. At that point, the spot that the cursor was over will be annotated with the specified label. When the expert user has a good idea about who or what is to be annotated, then the hotkeys can be put to work very efficiently. Instead of having to find the name desired and drag the name over to the position, the user can simply position the mouse and press the hotkey. This method is especially useful when a photo collection has only a few names that appear frequently.

Pop-up menu annotation (Figure 2c) also aims at reducing mouse movement in selecting labels to be used for annotation. This method offers a menu when the right mouse button is pressed on a picture. The menu consists of the currently selected labels in the list so that users can annotate pictures with a label in the menu at the position the right mouse button was pressed. This mechanism was revised again later as Label Paint Annotation so that a photo can be annotated with the currently selected labels whenever users click on a photo without selecting a pop-up menu.

In addition to the proposed methods for improving the speed in annotating individual photos, we also designed two additional methods for improving the efficiency of bulk annotation as follows:

- **Checkbox Annotation:** Select one or more photos and click the check box next to the label in the list (Figure 1).
- **Drag-Drop Annotation:** Select one or more photos and drag and drop the selected labels onto the photo. All the photos that were selected get annotated with the labels.

The best solution may be to include as many methods as possible while providing options for users to select and customize the methods. However it raises the question of how much the user can learn at first; if presented by too many options in the beginning, the user may become confused and frustrated. Therefore it may be optimal to use the multilayered approach (Kang, Plaisant, & Shneiderman, 2003, Shneiderman, 2003) when designing the initial interface.



(a) Split menu annotation



(b) Hotkey annotation



(c) Pop-up menu annotation

FIGURE 2 Enhanced direct annotation approaches within PhotoFinder.

2.3. Semiautomatic Annotation

The performance of direct annotation may be improved by supporting bulk annotation as described in the previous section. However this approach introduces two new challenges. First, unless photos to be annotated with the same labels are clustered together in some way so that they can be selected together, users need to manually select photos one by one before applying bulk annotation. Second, because multiple photos are annotated with a single interaction, the position of the labels in each photo cannot be explicitly specified by users.

To cope with these challenges, we explored a semiautomatic annotation strategy that takes advantage of human and computer strengths (Suh & Bederson, 2007). The semiautomatic approach enables users to efficiently update automatically obtained metadata interactively and incrementally. Even though automatically identified metadata are compromised with inaccurate recognition errors, the process of correcting inaccurate information can be faster and easier than manually adding new metadata from scratch. To facilitate efficient bulk annotation, two clustering algorithms were introduced to generate meaningful photo clusters (Suh & Bederson, 2007): *hierarchical event clustering* and *clothing-based person recognition*. The first method clusters photos based on their time stamps so that each group of photos corresponds to a meaningful event. The hierarchical event clustering identifies multiple event levels in personal photo collection and allows users to choose the right event granularity. The second method uses a clothing-based human model to group similar-looking people together. The clothing-based human model is based on the assumption that people who wear similar clothing and appear in photos taken within relatively short periods of time are very likely to be the same person.

To explore our semiautomatic strategies, we designed and implemented a prototype called SAPHARI (Semi-Automatic PHoto Annotation and Recognition Interface). The prototype provides an annotation framework focusing on making bulk annotations on automatically identified photo groups. The two automatic clustering algorithms provide meaningful clusters to users and make annotation more usable than relying solely on vision techniques such as face recognition or human identification in photos. In addition, the clothing-based person recognition automatically detects the positions of people in a photo, which can be used to position annotations. It is interesting that, although we found that absolute performance using this approach was not much better than manual annotation, the user study showed that the semiautomatic annotation was significantly preferred over manual annotation (Suh & Bederson, 2007). As computer vision techniques improve, the relative advantage of semiautomated techniques can only improve.

Figure 3 shows a prototype of SAPHARI, which shows the clusters of identified people whose upper bodies are cropped from photos of the same event. Because the automatic clustering techniques can also have errors, we designed the GUI so that users can correct errors manually by dragging a photo into the correct group. After error correction, users can annotate multiple photos at once (also with position information) by dragging a label onto the group.

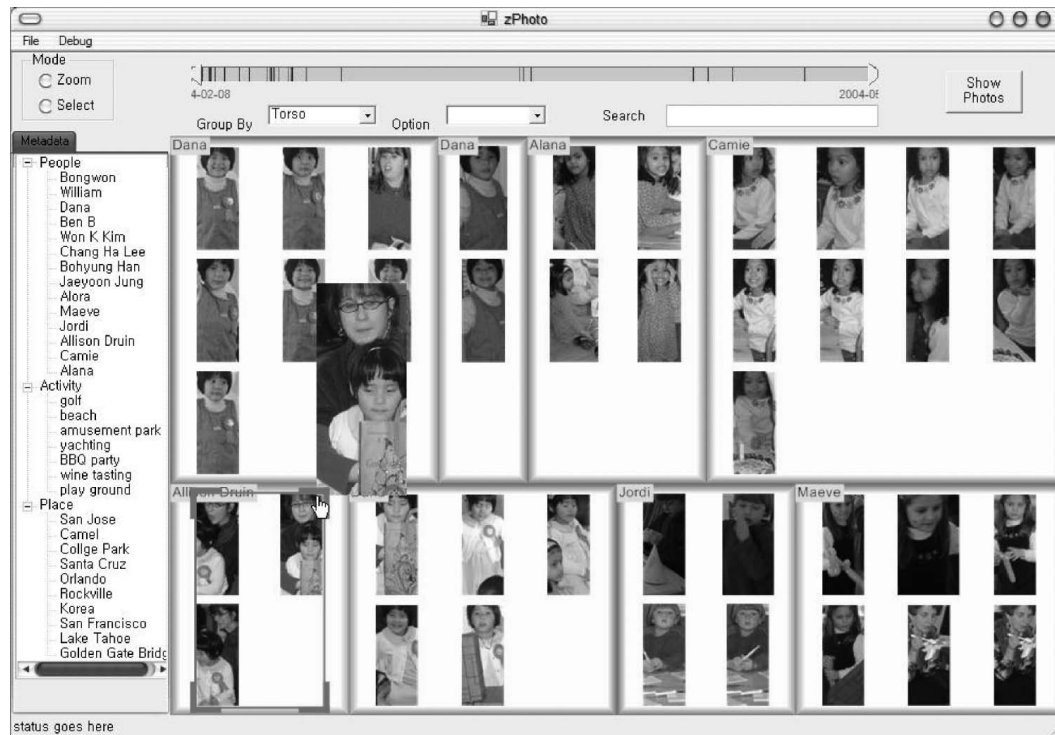


FIGURE 3 Identified people who are cropped from photos are laid out on the screen in the SAPHARI prototype. *Note.* Similar looking people are clustered based on their clothing, facilitating bulk annotation.

2.4. Image Annotation Discussion

There has been significant research on how to acquire useful metadata for images.

Many researchers have focused on automatic extraction of metadata by understanding images. For example, researchers have focused on automatic object detection such as face recognition and content-based categorization and retrieval (Yang, Kriegman, & Ahuja, 2002; Yoshitaka & Ichikawa, 1999; Zhao, Chellappa, Philips, & Rosenfeld, 2003; Chellappa, Wilson & Sirohey, 1995). However, such automatic techniques have achieved limited success so far when applied to personal photo management (Phillips et al., 2003; Suh & Bederson, 2007).

Rather than pure image-based approaches, some researchers used related context information to identify relevant metadata. For example, Naaman, Yeh, Garcia-Molina, and Paepcke (2005) used time and location information to generate label suggestions to identify the people in photos. Lieberman and Liu (2002) used relationships between text and photographs to semiautomatically annotate pictures.

On the other hand, Web-based collaborative approaches to collecting metadata become popular. Web pages, online photographs, and Web links have been actively annotated with rich metadata through tagging (e.g., Flickr and Del.icio.us). Tagging is a simple kind of annotation where keywords can be

associated with objects such as photos. However, the simple underlying approach belies the richness and power gained by broad communal access to these tags which enable cross-collection aggregation and searching along with innovative interfaces for creating and presenting those tags (i.e., tag clouds). With Flickr, users can share personal photos with friends, family, and colleagues, allowing the invited users to make annotations on shared photos. When users select photos and make them available to others, they seem to be willing to invest more effort in annotation (Shneiderman, Bederson, & Drucker, 2006). Also, by making them public, they invite others to comment and add annotations. We believe that this "folksonomy" based collaborative annotation has great potential for creating more useful metadata.

On the other hand, folksonomy based tagging system has a set of limitations. Folksonomic tagging can be inconsistent because of its vocabulary problem (Furnas, Landauer, Gomez, & Dumais, 1987). In folksonomy-based systems, there is no bona fide standard for selecting a tag, and users can choose to use any word as a tag. For example, one can use the tag *TV*, whereas others choose to use the tag *television*. Furthermore, Furnas et al. showed that it is not always easy for users to come up with good descriptive keywords that can be shared. Tags, therefore, can be easily idiosyncratic and often causes meta-noise, which decreases the retrieval efficiency of systems. Going further with collaborative annotation approaches, Von Ahn addressed the challenge by adding gaming to the mix. The ESP Game (Von Ahn & Dabbish, 2004) combines a leisure game with practical goals. It lets people play an image guessing game that gathers image metadata as a side effect of their playing.

Another ongoing challenge for metadata management is where to store the metadata. It is crucial that users own their metadata just as much as they own their photos. Yet some companies (such as Apple with their iPhoto software) create a custom database that separates metadata from the photos, leaving no easy way to get the annotations back for sharing or for use in other programs. Software developers tend to like centralized metadata stores because they make it easier to write software that perform fast searches (and because they "lock in" customers to their products). However this approach is not necessary. Instead, we suggest storing all annotations and metadata using the "IPTC" standard format (International Press Telecommunications Council; <http://www.iptc.org>) in the "EXIF" header (Japan Electronic and Information Technology Industries Association, 2002) of JPEG images. Then the application can create a cached copy of the metadata for efficient operations while leaving the "original" metadata with the photo. This is how we implemented PhotoMesa, and it appears to be the same approach taken by some other commercial software such as Picasa and ACDSee. However, even this approach has some challenges because of the following limitations:

- Not every image format supports metadata fields in the file headers.
- Metadata standards (i.e., EXIF/IPTC) do not have rich enough fields for objects such as people and hierarchical categories.
- Some operating systems make it difficult and possibly dangerous to modify the image header without changing the file modification date (which users sometimes use to determine when a file was created).

Although the issues and challenges relating to metadata standards are significant, they are beyond the scope of this article. We thus raised these few issues to point out the importance of getting it right, as it has such a dramatic affect on user experience. We therefore call on the industry to collaborate more openly in developing consistent and rich photo metadata standards addressing at least the aforementioned issues.

Technological change in society is rapid. There are new opportunities that we do not explore in this article. For example, the growing number of cell phones and even dedicated cameras that offer voice recording and network connectivity offer new possibilities such as supporting annotation at the moment the photo is captured. On the other hand, collaborative tagging has shown users are willing to make annotations when certain conditions are met. Further research is needed to investigate how the approaches we discuss in this article apply to those new settings.

Even though the importance of photo metadata is well known, the development of systems to support annotation has not been so successful. That is partly because of the limitations of automatic content analysis approaches, and partly because software designers have not understood in which situations users will be ready to spend a substantial amount of energy annotating photos. In this section, we have illustrated how novel interface designs might facilitate photo annotation by individuals or communities of users.

GUIs for using metadata in image management and exploration

Once richer image metadata is available either by automatic image content analysis or by manual annotation, users can make use of it for various image management tasks including not only search but also organization, meaning extraction, navigation, and distribution.

From our research experience with personal media management systems (Bederson, 2001; Bederson et al., 2002; Kang & Shneiderman, 2000, 2002, 2006; Khella & Bederson, 2004; Kules et al., 2004; Shneiderman et al., 2006; Shneiderman & Kang, 2000; Shneiderman et al., 2002; Suh & Bederson, 2007), we learned that managing personal media objects is a challenge for most users, who may struggle to understand, interpret, arrange, and use them. They wrestle with at least two major problems. First, most tools were designed and developed based on rigid and system-driven organizing metaphors (such as file-folder hierarchy). Second, those tools were not suitably designed to use metadata in exploring the media objects—beyond merely searching them. To address these challenges, we designed novel GUIs that hope to improve task performance as well as user satisfaction in exploring and managing personal media data.

3.1. Organization

With rich image metadata, even if users can find the images they need, they still frequently want to organize them for other reasons such as supporting future serendipitous browsing or to have the satisfaction of putting their images in order

(Balabanovic, Chu & Wolff, 2000; Kuchinsky et al., 1999; Rodden & Wood, 2003). Hence, supporting image organization is important.

One of the main challenges in designing a novel user interface for organizing images is how to let end-users represent and flexibly apply their conceptual models to their image collections. Users typically understand their data by constructing conceptual models in their minds. There is no unique or right model. Rather, the models are personal, have meaning for the individual who creates them, and are tied to specific tasks. Even in a simple personal photo library, images can be organized by time lines, locations, events, people, or other attributes, depending on users' conceptual models and specific tasks. Despite the diversity of users' conceptual models, the means available for users to organize and customize their information spaces are typically poor and driven mostly by storage and distribution models, not by users' needs.

Ben Shneiderman and Hyunmo Kang tried to resolve this challenge by providing users an environment to customize their information space appropriately for their conceptual models and specific tasks. We introduced a model called *Semantic Regions* (Kang & Shneiderman, 2006; see Figure 4), which are query regions drawn directly on a 2D information space. Users can specify the shapes, sizes, and positions of the regions and thus form the layout of the regions meaningful to them. Semantic Regions are spatially positioned and grouped on the 2D space based on personally defined clusters or well-known display representations such as a map, tree, time line, or organization chart. Once the regions are created and

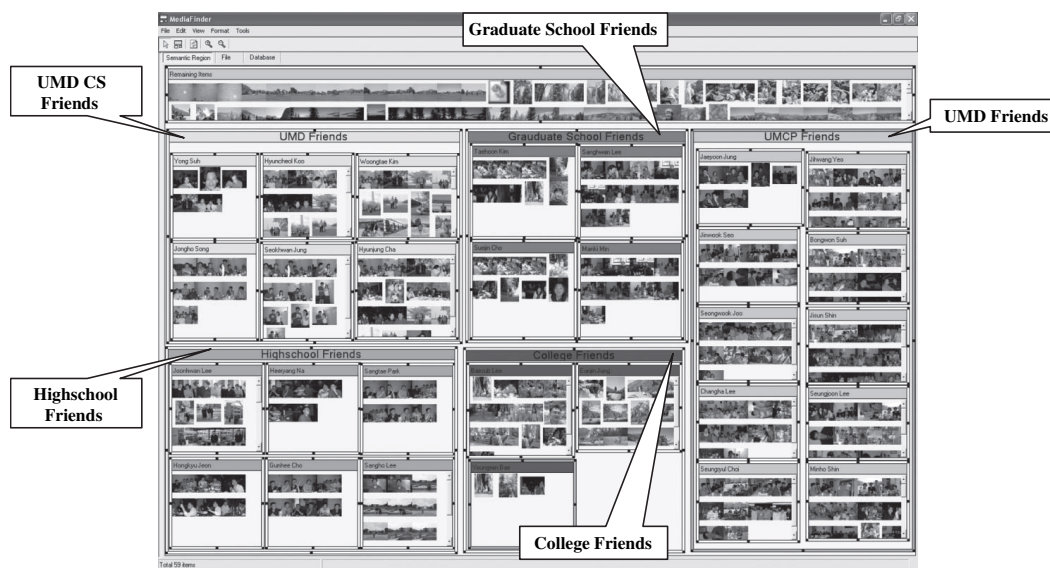


FIGURE 4 A friend group conceptual model: Each region represents a person and contains all the photos annotated with the name defined in it. *Note.* The regions are grouped into five clusters to represent different friend groups (UMD friends—University of Maryland computer science students-, high school friends, graduate school friends, college friends, and UMCP friends—University of Maryland noncomputer science students). Each group has its own color to represent the different group of friends.

arranged, users can specify semantics for each region using image metadata. The specified image metadata are conjunctively joined to form the semantics of a region.

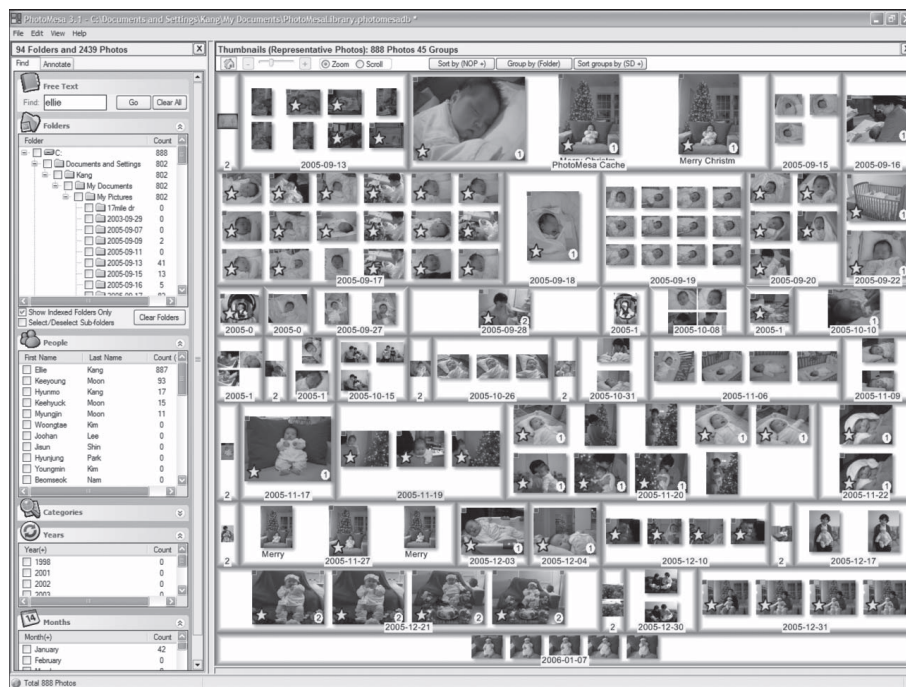
In Figure 4, each region represents a person, and the regions are grouped into five clusters to represent different friend groups. When the images are dragged onto the constructed model, they are automatically placed in the appropriate regions based on the annotations. This metaphor is called *fling-and-flock*; that is, users fling the objects and the objects flock to the regions. If photos do not satisfy any of the semantics of regions, they are collected in the *remaining items region* located at the top of the panel.

Small usability studies of the Semantic Region approach were conducted to discover potential user interface improvements and to gain a deeper understanding about users' ability to understand, construct, and use semantic regions. The studies mainly focused on qualitative attributes such as user satisfaction rather than task performance. From these studies (Kang & Shneiderman, 2006), we learned that users liked the idea of spatial organization and dynamic grouping, but they were not too interested in constructing their own models, because it took too much time and effort. Users wanted rich built-in templates rather than user-defined templates or more automatic features to reduce the steps for region construction and semantic specification. Based on these studies, we simplified this approach by removing steps from the region construction and semantic specification while supporting automatic grouping and layout based on image metadata.

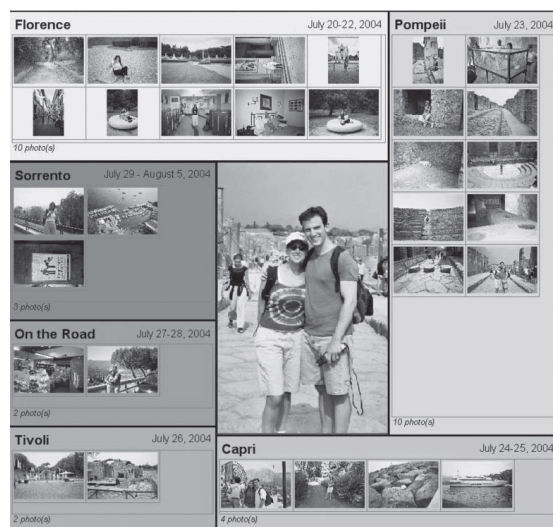
In a different project, we created an alternative and simplified approach for displaying groups of photos. This tool, called PhotoMesa, was developed as a successor to PhotoFinder by adding a zoomable user interface and the ability to show multiple groups of photos on the screen at the same time (Bederson, 2001; Bederson et al., 2002). In PhotoMesa, users can dynamically control how photos are grouped based on available metadata (Figure 5a). For example, photos can be grouped by the folder they belong to, by when the photos were taken (month or year), by people in the photo, or by any user-defined categories (locations, events, objects, etc.) that the photos are annotated with. The groups are automatically placed using quantum treemap space-filling algorithm (Bederson et al., 2002), and the groups can be sorted by photo metadata within a group such as number of photos in each group, starting date of photos in the group, title of the group, and so on. Each photo can be contained in multiple groups as long as it satisfies the grouping category.

Another approach created by Ben Shneiderman and Jack Kustanowitz is designed for groups of photos that have some relationship among them. In this case, it can be advantageous to show this relationship in the layout, rather than only in textual captions. Figure 5b shows a novel bilevel hierarchical layout, in which one region is designated for primary content, which can be a single photo, text, graphic, or combination (Kustanowitz & Shneiderman, 2005, 2006). Adjacent to that primary region, groups of photos are placed radially in an ordered fashion, such that the relationship of the single primary region to its many secondary regions is immediately apparent. A compelling aspect is the interactive experience in which the layout is dynamically resizable, allowing users to rapidly, incrementally, and reversibly alter the dimensions and content.

(a)



(b)



(c)

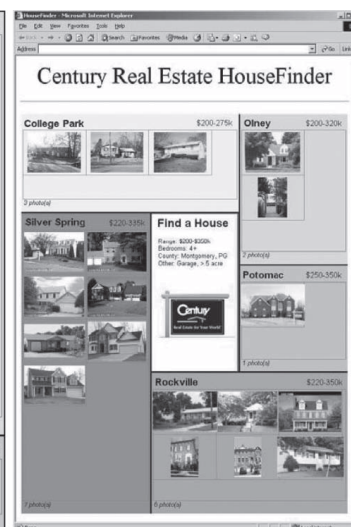


FIGURE 5 Organization of personal photo collections. (a) PhotoMesa interface enables users to dynamically group photos based on photo metadata. The group layout is automatically generated using space filling algorithm but they can also be sorted by the metadata of group photos. PhotoMesa enables users to control the visible photos in various ways. In this example, PhotoMesa shows only the representative photos of each group. (b) Bi-level hierarchical layout shows the photo groups taken in 6 locations during a trip by placing the groups clockwise in chronological order (left). (c) A real estate browser with 5 communities and the price ranges of selected homes is shown on the right.

3.2. Navigation and Browsing

Image browsing is important for a number of reasons. First, no matter what information retrieval system is being used, the user has to browse the results of the search. It is certainly important to build query systems that help users get results that are as close to what is wanted as possible. However, there will always be images that need to be browsed visually to make the final pick. The second reason is that sometimes people browse images just for the pleasure of looking at those images, and they often do it with other people. This is especially true for personal photos. Looking at home photos has a lot of overlap with traditional retrieval systems. People still want to be able to find photos of particular people, events, and so on. However, they are less likely to be time pressured to find a particular photo and more likely to be interested in serendipity (i.e., finding photos they weren't looking for; Balabanovic, Chu, & Wolff, 2000; Bederson, 2001; Rodden & Wood, 2003, Shneiderman et al., 2006).

We designed PhotoMesa to support this serendipitous style of use. It uses a zoomable user interface (Figure 6) to enable users to see all the images in a single view and concentrate on the images while browsing photos by zooming in or out without the need for interacting with other UI widgets such as scrollbar or pop-up window. This design is also based on the characteristics of personal image collections. As users have seen them at least once, they already have rough idea about their images and metadata. Hence, they often want to start browsing from the overview of their whole collection without worrying about missing some photos by accident (because they scrolled outside the currently visible region).

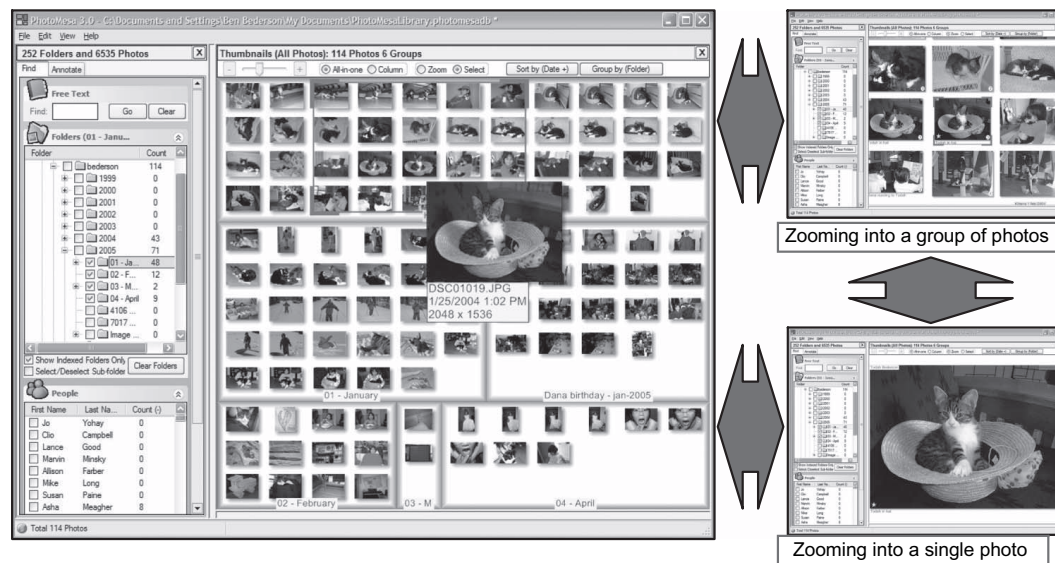


FIGURE 6 PhotoMesa shows all photos in a single view. *Note.* The layout of photos depends on how many groups and photos have been loaded. A larger preview of any photo can be obtained by moving the mouse over one of the small thumbnails. PhotoMesa uses a Zoomable User Interface and supports browsing of photos by zooming in or out. The order as well as the group that photos are displayed in can be controlled dynamically by photo metadata.

This zoomable browsing approach was further enhanced in terms of system performance and users' cognitive workload by supporting semantic zooming (Bederson et al., 1996). For example, if there are too many photos for users to recognize any of them, PhotoMesa has the option of displaying just representative photos (Figure 4a). *Time quilt* (Huynh, Drucker, Baudisch, & Wong, 2005; see Figure 7a) showed the improvement in photo browsing performance by making use of a linear timeline on top of the 2D space-filling techniques with semantic zooming. This layout is designed to convey a temporal order while making reasonable use of screen space by packing a time line layout into a rectangular screen space using a "line break" algorithm. Their user study shows that this combined layout enabled users to complete some photo browsing tasks faster than the quantum treemap layout and the time line layout. A new tool called PhotoSynth (Snaveley, Seitz, & Szeliski, 2006; Figure 7b) uses 3D space for photo browsing. The system takes a large collection of photos of a place or an object, analyzes the photos for similarities, and displays them in a reconstructed 3D space so that users can interactively browse and explore large collections of photos of a scene using a 3D interface.

In addition to the design of novel browsing interfaces, we also developed a novel thumbnail generation algorithm that aims to reduce users' cognitive workload during browsing (Suh, Ling, Bederson, & Jacobs, 2003). Using thumbnails has a trade-off; bigger thumbnails use more screen space, but smaller thumbnails are hard to recognize. Automatic thumbnail cropping is a method to find the optimal part of an image for thumbnail generation. The algorithm finds the informative portion of images based on a saliency model and cuts out the nonsalient part of images (Figure 8). Thumbnail images generated from the cropped part of images increases users' recognition and helps users in visual search. On the other hand, the face detection based cropping technique (Suh et al., 2003) shows how semantic information can be used to enhance thumbnail generation further (when that semantic information about an image is known). For example, using face detection can more effectively create thumbnails with the images of people, which was demonstrated to significantly increase users' recognition and visual search performance.

3.3. Search and Meaning Extraction

There is no doubt that search is a hugely important photo management task. Users may be looking for photos of their grandmother, their trip to Disney World, or a friend's wedding. Interactive visualization to select and view dozens or hundreds of photos extracted from tens of thousands has become a popular strategy (Shneiderman et al., 2006). In addition to simple photo search based on annotation, users sometimes want to extract meaning from their photo collections such as, Who else appeared in the pictures of my grandmother and me over 20 years and when and where did I meet her? Who appeared both in my friend's wedding pictures and in my wife's friend's wedding pictures? How many photos did I take of my daughter each year? What were the special events since she was born? Any seasonal differences in photo distribution. This kind of meaning extraction from personal photo collections becomes

(a)



(b)



FIGURE 7 Photo browsing interfaces. (a) Time quilt—a layout designed to convey temporal order while making better use of screenspace than a timeline, showing approximately 5,500 photos with representative thumbnail overview. (b) PhotoSynth takes a large collection of photos of a place or an object, analyzes them for similarities, and displays them in a reconstructed three-dimensional space.

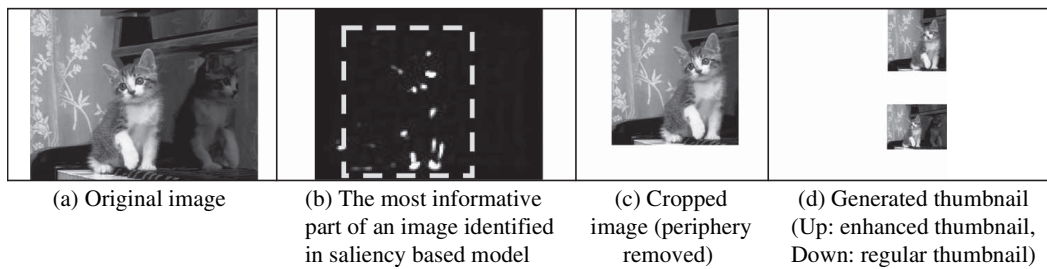


FIGURE 8 The saliency based model identifies the most informative part of an image. Thumbnail images created out of the core part help users' recognition and visual search performance.

more common and popular tasks as users have an increasing number of digital photos and richer metadata. However, little work has been done so far in this area.

A related effort in this area is the Semantic Region project (Kang & Shneiderman, 2006) described previously. It provides an interactive visualization technique called *region brushing* to support meaning extraction by visualizing the relationships among the semantic regions. Region brushing is used to highlight the personal media items contained in multiple regions simultaneously (Figure 9a) and is often used for depicting both the intrarelations and the interrelationships among models. Because a personal media item can be contained in multiple semantic regions, the existence as well as the amount of the shared personal media items among the regions represents their relationships. In Figure 9b, three models; HCI researcher, U.S. map, and CHI conference calendar are combined together. Because a photo can be contained in multiple regions across the different models; many questions concerning the interrelationships of models can be

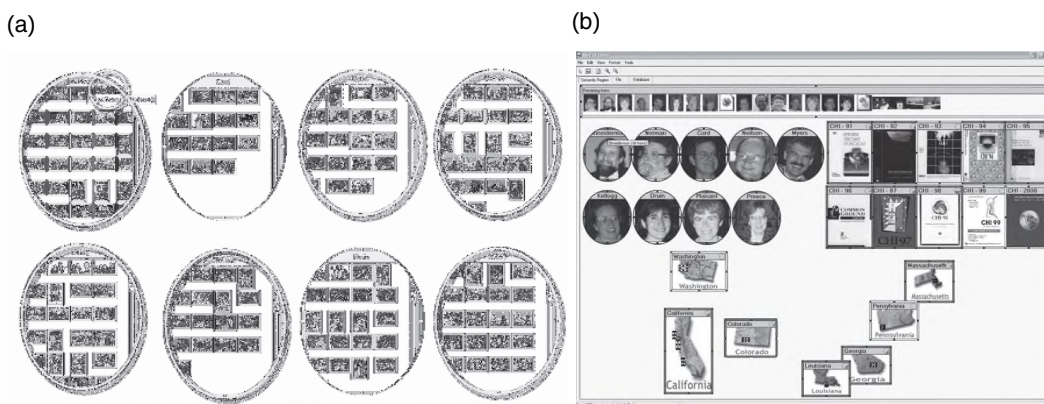


FIGURE 9 Region brushing: An interaction technique to visualize the shared items among the Semantic Regions. *Note.* (a) regions containing any of the photos in the region the mouse is over (Shneiderman's region) are highlighted. (b) visualization of interrelationships of the multiple models. By placing the mouse on Shneiderman's region, users can see which conference and which state his photos were taken.

answered through region brushing. Such questions might be, “Find the name of the conferences that all nine people participated in and where was the conference held?” “What was the name of the conference held in Atlanta, Georgia, and who did not appear in the photos taken at this conference?” “Find the conferences in which Ben Shneiderman did not appear in the photos and where was the conference held?” and so on. Furthermore, constructing new semantic regions recursively inside a region (e.g., location regions inside a person region) enables users to dynamically regroup the photos contained in a region and extract meanings.

Techniques such as *query preview* (Greene et al., 1999) or *faceted search* (Yee, Swearingen, Li, & Hearst, 2003) can help users extract meanings as well as rapidly filter down results sets by showing counts of available items along multiple dimensions or facets. PhotoMesa applies these approaches by letting users search for photos using the Find tab (left panel in Figure 10) in six ways: with keywords, by the folder a photo is in, by the people in a photo, by the categories a photo was annotated with, by the year a photo was taken, or by the month a photo was taken. If a new photo set is searched by combining those six ways, each search panel previews the distribution of the searched photos so that users can see how

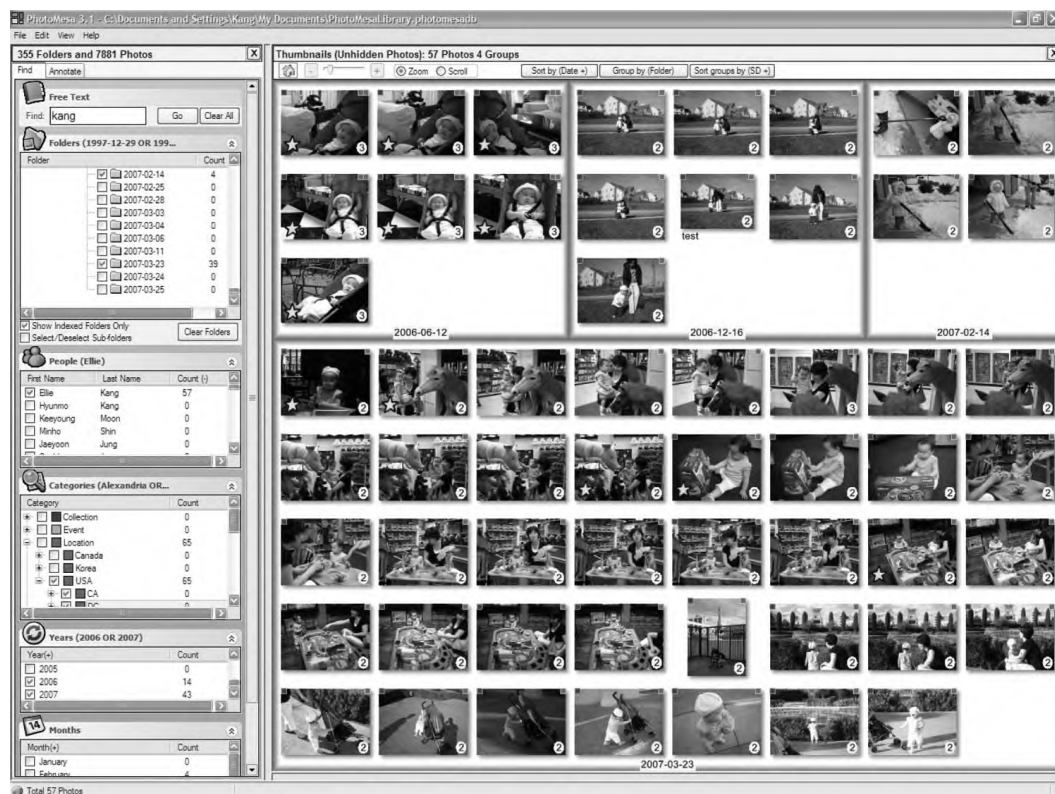


FIGURE 10 PhotoMesa search panel on the left enables users to search photos in six ways: with keywords; by the folder a photo is in; by the people in a photo; by the categories a photo was annotated with; and by the year or month a photo was taken. *Note.* The searched photos are displayed in the thumbnail viewer on the right and their distributions are shown in each search panel.

many photos are from each folder, how many people (and who) appear in the searched photos, where the searched photos were taken (or how many photos were from other user-defined categories), and when they were taken (month and year). With this feature, users can get information about the searched photos and use it for further exploratory search.

3.4. Sharing and Distribution

Photo sharing is becoming one of the most common activities of digital photo management. Our earlier work provided interesting insights about how individuals and communities share their personal photographs. For example, a client-server model of PhotoFinder, PhotoFinder Kiosk (which was installed at CHI 2001), enabled members of a community of HCI professionals to enjoy, share, and annotate photos at an interactive public exhibit (Kules et al., 2004; Shneiderman et al., 2002). There is now a tremendous commercial activity in this area, with numerous popular photo sharing Web sites such as Flickr and Picasa Web Albums.

We previously mentioned some of the trade-offs of desktop applications and Web-based photo-sharing sites. However, the current bifurcation requires users to pick one approach or the other to manage their photos. We believe that this split is an artifact of business requirements and software development practices and that with a little more creativity and effort, it is possible to build a hybrid solution that bridges both worlds providing the advantages of each. For example, Sharpcast (<http://www.sharpcast.com>; Sharpcast) enables users to share, access, and organize their photos through either local file systems or Web sites. We also have heard of people using independent file synchronization tools, such as FolderShare (<http://www.foldershare.com>; Microsoft) to keep multiple copies of their photos synched with metadata between different machines, including Web servers. All this points to the necessity of a hybrid approach. From a user's perspective, it doesn't make any sense that they should have to choose in advance whether they should commit to storing photos locally or online.

To assess the usefulness of the hybrid approach, we built a prototype extension to PhotoMesa that hides the distinction between local and Web access from the user. Because photos almost always start locally, after being transferred from a camera to a computer, we assume users will open the photos regularly in their local tool. Then the tool continuously synchronizes the local photos with the Flickr Web site using its public API in the background. In this manner, all the user's photos are also available within their private Flickr account. Furthermore, any changes or annotations to the photos on Flickr are synched back to the desktop application the next time they are accessed locally. In addition, full searches of Flickr's remote database can be performed within PhotoMesa the same way local searches are made, and remote photos are shown grouped by whatever metadata Flickr makes available.

There is a largely unmet need of users to be freed from the shackles of tightly controlled applications that limit access to their photos and metadata. In the increasingly decentralized world of information, that iron-fisted approach is just not sustainable anymore.

4. CONCLUSION

The world of digital photo management has come a long way in the past few years, and we are pleased to see that the basic needs of most personal photo users are now being met. However we aren't done yet. There remain large gaps in meeting users' needs such as recording richer annotations that are more accurate and broadly available. It is crucial to support seamless sharing of photos and their stories in a way that lets multiple parties participate without losing ownership of the photos and metadata.

It is not clear how well current approaches will scale up. With the ever-shrinking cost of storage at the same time that cameras are becoming ubiquitous, we can expect that over an individual's life, it will be common to manage tens or even hundreds of thousands of photos.

The challenge of applying automated or human-assisted analysis techniques to these photos in a way that remains understandable and usable is ongoing. It also remains a challenge to find the right balance between local and Web access that enables people to have reliable long-term access to their full resolution photos in a way that does not lock them into any particular vendor's system. People must also be able to maintain their privacy and control over photos while sharing and collaboratively editing the photos and stories about them. When we envision the addition of audio and video to these collections, which is already happening, all of these challenges grow even further.

The transition to digital photography has already largely happened. This implies that consumers are satisfied enough with current solutions—or at least prefer digital to film. However that doesn't mean that their long-term problems are solved or that things are good as they could and should be. It is our job as designers and developers of these systems to address these challenges.

In this article, we described a range of projects that attempt to improve some of the problems already described. By integrating better annotation tools, visual presentation, and query systems, and through the use of more open metadata storage approaches, we have the opportunity to take the growing digital photo market and accelerate improvements for all of us to use with our own dear personal photos. Some directions not covered in this article, such as integrating annotations into other life activities (such as blogging, diary writing, e-mail, etc.) also offer promise. In the end, we expect a rich ecosystem of photo-related technologies will continue to evolve that support the diverse set of users that take photos.

REFERENCES

- Balabanovic, M., Chu, L. L., & Wolff, G. J. (2000). Storytelling with digital photographs. In *Proceedings of Human Factors in Computing Systems* (pp. 564–571). New York: ACM Press.
- Bederson, B. B., Hollan, J. D., Perlin, K., Meyer, J., Bacon, D., & Furnas, G. W. (1996). Pad++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics. *Journal of Visual Languages and Computing*, 7, pp. 3–31.
- Bederson, B. B. (2001). PhotoMesa: A zoomable image browser using quantum treemaps and bubblemaps. In *Proceedings of ACM Conference on User Interface and Software Technology* (pp. 71–80). New York: ACM Press.

- Bederson, B. B., Shneiderman, B., & Wattenberg, M. (2002). Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Transactions on Graphics*, 21, 833–854.
- Chalfen, R. (1987). *Snapshot versions of life*. Bowling Green, OH: Bowling Green State University Popular Press.
- Chellappa, R., Wilson, C. L., & Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 705–740.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communication of ACM*, 30, 964–971.
- Goto, Y., Jung, J., Ma, K., & McCaslin, O. (2000). *The effect of direct annotation on speed and satisfaction*. Student HCI Online Research Experiment, Department of Computer Science, University of Maryland, College Park, MD. Retrieved from <http://www.otal.umd.edu/SHORE2000/annotation/>
- Greene, S., Tanin, E., Plaisant, C., Shneiderman, B., Olsen, L., Major, G., et al. (1999). The end of zero-hit queries: Query previews for NASA's Global Change Master Directory. *International Journal of Digital Libraries*, 2(2–3), 79–90.
- Huynh, D., Drucker, S., Baudisch, P., & Wong, C. (2005). Time quilt: Scaling up zoomable photo browsers for large, unstructured photo collections. *Extended Abstracts of Human Factors in Computing Systems*, 1937–1940.
- Japan Electronic and Information Technology Industries Association. (2002). Exchangeable image file format for digital still cameras: Exif Version 2.2. Retrieved from <http://www.exif.org>
- Jung, J. (2000). Empirical comparison of four accelerators for direct annotation of photos. Retrieved from <http://www.cs.umd.edu/hcil/photolib/paper/cmsc498paper.doc>
- Kang, H., Plaisant, C., & Shneiderman, B. (2003). New approaches to help users get started with visual interfaces: Multi-layered interfaces and integrated initial guidance. *Proceedings of Digital Government Conference*, 1–6.
- Kang, H., & Shneiderman, B. (2000). Visualization methods for personal photo collections: Browsing and searching in the PhotoFinder. *Proceedings of IEEE International Conference on Multimedia and Expo*, 1539–1542.
- Kang, H., & Shneiderman, B. (2002). Dynamic layout management in a multimedia bulletin board. *Proceeding of IEEE International Symposium on Human-Centric Computing Languages and Environments*, 51–53.
- Kang, H., & Shneiderman, B. (2006). Exploring personal media: A spatial interface supporting user-defined semantic regions. *Journal of Visual Language and Computing*, 17, 254–283.
- Khella, A., & Bederson, B. B. (2004). Pocket PhotoMesa: A zoomable image browser for PDAs. In *Proceedings of Mobile and Ubiquitous Multimedia* (pp. 19–24). New York: ACM Press.
- Kuchinsky, A., Pering, C., Creech, M. L., Freeze, D., Serra, B., & Gwizdka, J. (1999). FotoFile: A consumer multimedia organization and retrieval system. *Proceedings of ACM Conference on Human Factors in Computing Systems*, 496–503.
- Kules, B., Kang, H., Plaisant, P., Rose, A., & Shneiderman, B. (2004). Immediate usability: A case study of public access design for a community photo library. *Journal of Interacting With Computers*, 16, 1171–1193.
- Kustanowitz, J., & Shneiderman, B. (2005). Meaningful presentations of photo libraries: Rationale and applications of bi-level radial quantum layouts. In *Proceedings of Joint Conference on Digital Libraries* (pp. 188–196). New York: ACM Press.
- Kustanowitz, J., & Shneiderman, B. (2006). Hierarchical layouts for photo libraries. *IEEE MultiMedia*, 13, 62–72.
- Lieberman, H., & Liu, H. (2002). Adaptive linking between text and photos using common sense reasoning. *Proceedings of Adaptive Hypermedia and Adaptive Web Systems*, 2–11.

- Naaman, M. (2005). Leveraging geo-referenced digital Photographs. Stanford University, Stanford, CA.
- Naaman, M., Yeh, R. B., Garcia-Molina, H., & Paepcke, A. (2005). Leveraging context to resolve identity in photo albums. *Proceedings of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries*, 178–187.
- Phillips, P. J., Grother, P. J., Michaels, R. J., Blackburn, D. M., Tabassi, E., & Bone, J. M. (2003). Face recognition vendor test 2002: Evaluation report, NISTIR 6965. Retrieved from <http://www.frvt.org/FRVT2002/default.htm>
- Rodden, K., & Wood, K. (2003). How do people manage their digital photographs? *Proceedings of ACM Conference on Human Factors in Computing Systems*, 409–416.
- Sears, A., & Shneiderman, B. (1994). Split menus: Effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction*, 1, 27–51.
- Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *IEEE Computer*, 16(8), 57–69.
- Shneiderman, B. (2003). Promoting universal usability with multi-layer interface design. In *ACM Conference on Universal Usability* (pp. 1–8). New York: ACM Press.
- Shneiderman, B., Bederson, B. B., & Drucker, S. M. (2006). Find that photo!: Interface strategies to annotate, browse, and share. *Communications of the ACM*, 49, 69–71.
- Shneiderman, B., & Kang, H. (2000). Direct annotation: A drag-and-drop strategy for labeling photos. *Proceedings of International Conference Information Visualization (IV2000)*, 88–95.
- Shneiderman, B., Kang, H., Kules, B., Plaisant, C., Rose, A., & Rucheir, R. (2002). A photo history of SIGCHI: Evolution of design from personal to public. *ACM Interactions*, 9(3), 17–23.
- Snavey, N., Seitz, S. M., & Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3), 835–846.
- Suh, B., & Bederson, B. B. (2007). Semi-automatic image annotation strategies using event based clustering and clothing based person recognition. *Journal of Interacting with Computers* 19(4), 524–544.
- Suh, B., Ling, H., Bederson, B. B., & Jacobs, D. (2003). Automatic thumbnail cropping and its effectiveness. *Proceedings of ACM Conference on User Interface and Software Technology*, 95–104.
- Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of ACM Conference on Human Factors in Computing Systems* (pp. 319–326). New York: ACM Press.
- Yang, M., Kriegman, D., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34–58.
- Yee, K-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. In *Proceedings of ACM Conference on Human Factors in Computing Systems* (pp. 401–408). New York: ACM Press.
- Yoshitaka, A., & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1), 81–93.
- Zhao, W., Chellappa, R., Philips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399–458.

Capture, Annotate, Browse, Find, Share:

Novel Interfaces for Personal Photo Management

Hyunmo Kang¹, Benjamin B. Bederson^{1,2}

Human-Computer Interaction Laboratory

University of Maryland Institute

for Advanced Computer Studies¹

Computer Science Department²

College Park, MD 20742

{kang, bederson}@cs.umd.edu

Bongwon Suh

Palo Alto Research Center

3333 Coyote Hill Road

Palo Alto, CA 94304

suh@parc.com

Abstract

The vision of ubiquitous digital photos has arrived. Yet, despite their broad popularity, significant shortfalls remain in the tools used to manage them. We believe that with a bit more creativity and effort, the photo industry can solve many of these problems, offering tools which better support accurate, rapid, and safe shared annotations with comfortable and efficient browsing and search. In this article, we review a number of projects of ours and others which relate to Ben Shneiderman's work on interfaces for photo management. With these, we describe the problems that we see in existing tools and our vision for improving them.

1. Introduction

The days when people debated the relative merits of film vs. digital imagery now seem almost quaint. And with hindsight, film seems destined to have been a blip in history along with LP vinyl records – a temporary physical analog recording medium. While some aficionados may still prefer some qualities of those dust-gathering mediums, the advantages of digital media have become clear. The rapid and inexpensive ability to edit, annotate, search, share and access has brought digital media to ubiquity.

And yet, with all this promise, shortfalls remain in the overall user experience. How many of us have replaced old unlooked at shoeboxes of prints with unlooked at digital archives of image files? How much has our ability to find a particular set of photos really improved (i.e., can you find those photos of a visiting uncle with your sister when they were children?) How do we record stories from our parents describing family photos? And how do we make sure those stories stay with the photo in question and get distributed to all copies of that photo within the family? And perhaps of most concern, how do we insure that these annotations stand the test of time and remain accessible as computers, file formats, recording mediums, and software change?

These changes are all happening within the context of human behavior, which does not change so rapidly. People still like immediate gratification, take pictures rapidly, print them, and share in social settings. Some people spend a lot of effort creating photo albums or “scrapbooking”. And of course, many do not. Understanding which behaviors

are fundamental and which are side effects of current technology is crucial because this understanding can and should influence where researchers spend their effort.

We explore these issues and more with much of the intellectual motivation coming from Ben Shneiderman, our close colleague who has pushed for a deeper understanding of and better support of photo management for over 10 years. His personal photo archives document the field of HCI going back to its beginning. He regularly shares those photos with great enthusiasm to visitors, motivating and exciting all of us – largely because of the care and consistency he has applied to annotating and organizing his photos. He regularly pulls up old photos of lab visitors showing everyone what they worked on 5 or 15 years ago (and of course, showing what they looked like too!). His early exploration of tools to support photo management (with co-author Kang) led to PhotoFinder [Kang and Shneiderman, 2000], and the ensuing PhotoMesa tools [Bederson, 2001; Bederson et al., 2002]. His personal interest helped inspire the authors of this article as well as other lab members (particularly Catherine Plaisant) to pursue the development of approaches and software to improve all of our user experiences when managing photos.

This, of course, all happened during a time of tremendous commercial activity in this area. There are wildly popular photo sharing websites, such as Flickr¹, Picasa Web Albums², Snapfish³, Shutterfly⁴, and PhotoBucket⁵ as well as equally well-used desktop photo

¹ Yahoo! Inc. www.flickr.com

² Google Inc. picasaweb.google.com

³ HP Inc., www.snapfish.com

⁴ Shutterfly Inc., www.shutterfly.com

⁵ Photobucket Inc. www.photobucket.com

applications such as Google Picasa⁶, Adobe PhotoShop Album⁷, and ACDSee. The two approaches (desktop application and website) are interesting to look at because they each offer distinct advantages to users. For example, websites are available anywhere and facilitate sharing, while desktop applications are faster, support higher resolution photos more easily, provide local ownership of photos, and offer richer interaction capabilities. Interestingly, each approach is gaining characteristics of the other. Web applications begin to offer dynamic, interactive content rather than static html pages through AJAX and Flash technologies. In addition, they often include plugins to ease uploading or improve performance and some offer APIs to enable desktop applications to access their data directly. At the same time, many desktop applications are offering web capabilities such as sharing.

Yet even with this commercial activity, the full potential of personal photo management has not been reached. There is the opportunity for richer annotation interfaces, automated content analysis, improved sharing and more creative organizational strategies. Our hope is that more photos end up with better metadata, enabling faster, easier and more accurate and enjoyable retrieval and use.

In this article, we look at some of the key activities and behavior patterns of personal photo users and examine how innovative user interfaces have the potential to enhance users' power, satisfaction, and control in managing and exploring their images. Starting with a close look at annotation, we examine how a combination of manual and automated

⁶ Google Inc. www.picasa.com

⁷ Adobe Systems Inc. www.adobe.com/products/photoshopalbum/starter.html

techniques can improve how people associate metadata with photos. We then look at how the resulting richer metadata can enable better interfaces for searching and browsing photos. Finally, we end with a discussion about the importance of sharing photos and how new interfaces enable that.

2. GUIs for Annotation

An essential question is how valuable photo metadata is. Our own assessment of user needs [Shneiderman and Kang, 2000] coupled with reports from other researchers [Chalfen, 1987; Naaman, 2005; Rodden and Wood, 2003], and our personal experience come together on this. They indicate that the photo metadata such as dates, locations, and content of the photos (especially people's names) play a crucial role in management and retrieval of personal photo collections.

However, in spite of the high utility of the photo metadata, the usability of software for acquiring and managing the metadata has been poor. The manual photo annotation techniques typically supported by photo software are time-consuming, tedious and error-prone, and users typically put little effort into annotating their photos. In fact, the industry attitude tends to be that since users do not annotate their photos very much, it is not necessary to spend much energy adding good support for it. However the success of keyword labeling, so called *tagging*, systems such as del.icio.us **Error! Bookmark not defined.** and Flickr **Error! Bookmark not defined.** hints that users indeed want to make annotations when reasonable utility and usability are supported. We believe that the photo annotation has enough utility for some users and it is the usability of software that needs to be improved.

In this section, a few innovative approaches will be presented to show how interaction and GUI design can improve the usability of the photo annotation process when they are

based on careful understanding of users' behavior and usage pattern. In addition, we will explain how those designs have been evolving over time to support a broader range of annotation tasks by combining the accessible technologies with the analysis of users and their needs.

2.1 Advanced Manual Annotation: Direct Annotation

Since annotations based on automated image content analysis are still limited, we developed an advanced manual annotation mechanism that can significantly reduce users' annotation workload under certain circumstances. From the observations of personal photo annotations [Shneiderman and Kang, 2000], we found that there were three interesting characteristics that could be useful for our interaction design.

- Personal photo libraries often contain many images of the same people at different events. In the libraries we looked at, we typically found 100-200 identifiable people in several thousand photos. Furthermore, the library has a highly skewed distribution with immediate family members and close friends appearing very frequently.
- Textual search often doesn't work reliably because of inconsistency in names with misspellings or variants (e.g. Bill, Billy, William).
- Lists of names of people appearing in photos are often difficult to associate with individuals, especially in group shots. Textual captions often indicate left-to-right

ordering in front and back rows, or give even more specific identification of who is where. However this approach is tedious and error-prone.

Based on these observations, we collaborated with Ben Shneiderman to develop the concept of *direct annotation*⁸ which is a technique using selectable, draggable labels that can be placed directly on the photo [Shneiderman and Kang, 2000]. Similar interfaces have also appeared recently on Flickr and MySpace. Users can select from a scrolling or pop-up list and drag by mouse or touch screen. This applies direct manipulation principles [Shneiderman, 1983, 2005] that avoid the use of a keyboard, except to enter a name the first time it appears. The name labels can be moved or hidden, and their presence is recorded in the database or in the header of an image file in a resolution-independent manner. The relative location of the target is stored based on an origin in the upper left hand corner of the photo with the point in the range (0, 0) – (1.0, 1.0) corresponding to the full image. This approach not only associates a name with a position on the photo, but also ensures that each name is always spelled the same way.

This simple rapid process also enables users to annotate at any time. They can add annotations when they first see their photos on the screen, when they review them and make selections, or when they are showing them to others. This design, which supports continuous annotation encouraged users to do more annotation, especially in collaborative situations such as with PhotoFinder Kiosk [Kules et al., 2004, Shneiderman et al., 2002]. In a public setting, visitors of PhotoFinder Kiosk added 1,335 name labels using direct annotation while adding 399 captions using traditional type-in method.

⁸ US Patent #7010751

The direct annotation mechanism was revised later so that it enables users to define their own categories such as activities, events, locations, objects in a photo in a hierarchical way (Figure 1) in addition to person names. A few alternative and complementary direct annotation mechanisms such as split menu annotation, hotkey annotation, popup menu annotation have also been designed and developed to accelerate the annotation process. These are described in more detail in the following subsection.

An informal experiment was conducted to see if the direct annotation method improved the annotation process in terms of annotation time and users' subjective preference compared with the traditional caption method or the click-and-type method [Goto et al, 2000][Jung 2000]. 48 volunteers participated in the experiment and a within-subject design was used, whereby each subject attempted an annotation task (20 names in five photos) on each system. While no significant difference was found in the mean annotation time, the direct annotation method was significantly preferred. In addition, there was a trend towards faster annotations with direct annotation. We believe that a more formal follow-up user study is required to identify under what circumstances (e.g. number of total labels, average number of people in a picture, pattern of people's appearance in a personal photo collection, and so on) the direct annotation mechanism works better than other annotation mechanisms in terms of several dependent variables such as completion time, error rate, users' satisfaction and confidence.

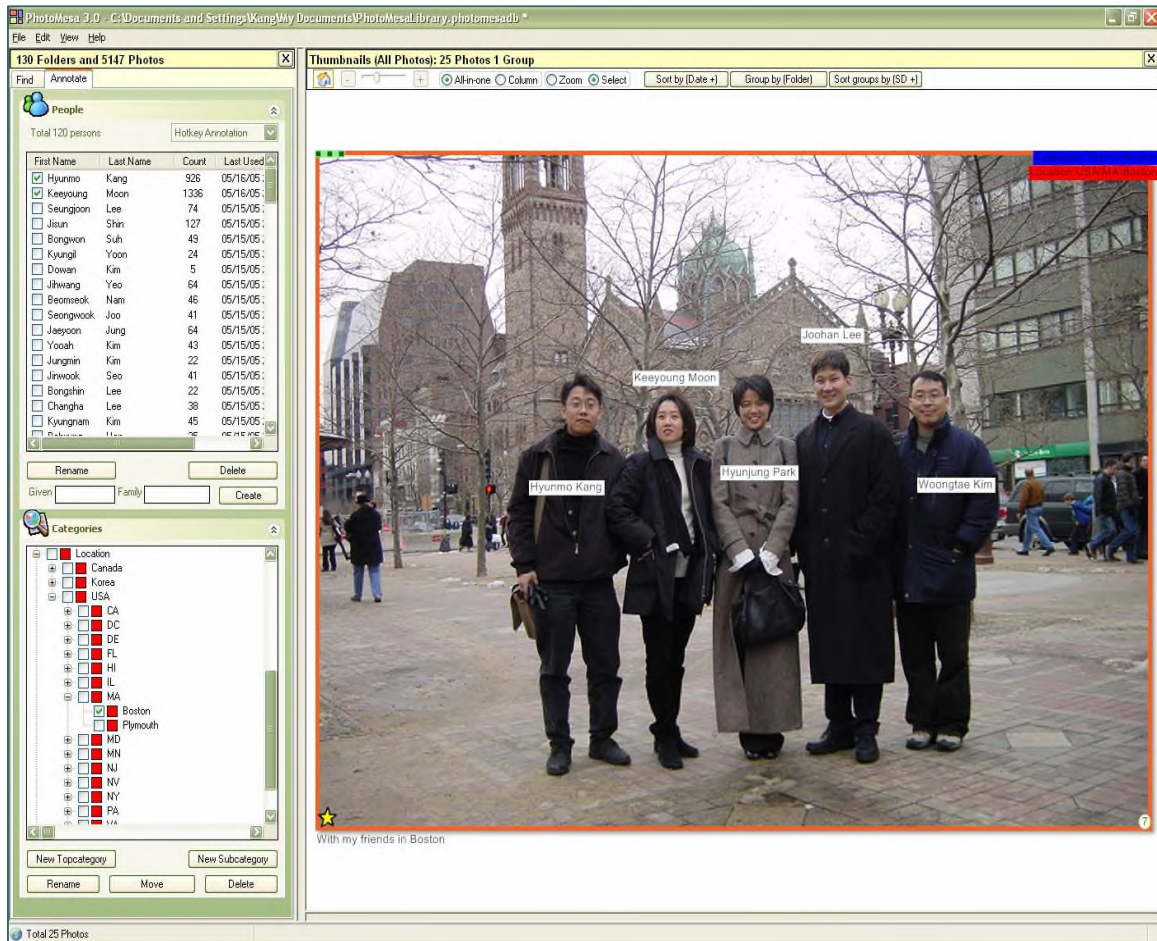


Figure 1: The revised direct annotation mechanisms as implemented in PhotoMesa, the successor to PhotoFinder. Users can add a caption under a photo, or add a particular attribute such as favorite (a yellow star on the bottom left of photo) or hidden. In addition, labels can be dragged from the list of people (on the top left) or from the user-defined category tree (on the bottom left). A label can be directly placed on the photo to represent where the individuals or objects are located within the photo.

2.2 Enhanced Direct Annotation: Bulk Annotation

The direct annotation mechanism was enhanced through a series of design cycles to support more efficient and productive annotation. Perhaps the most notable one was applying the "bulk annotation" technique, which lets users annotate several photos with a single interaction [Kuchinsky et al., 1999]. Bulk annotation is especially useful when users have a set of photos with the same person or group of people involved in the same

events. We thus designed and developed several variations of direct annotation and bulk annotation mechanisms to accelerate the annotation process.

The split menu annotation mechanism (Figure 2(a)) was designed to minimize users' time of scrolling the list to find correct labels to be used for photo annotation. The split menu featured a resizable splitter bar so that the number of the most frequent labels displayed in the top window. The scrollbar was removed from the top window, while the bottom half retained its scrollbar. The split menu raises interesting questions about what kind of automatic algorithms should be used to predict users' future access and facilitate rapid annotation.

Because some annotations get used so much more frequently than others, we designed the interface so that each label (either person's name or categories) can be associated with a "hot key" (Figure 2(b)). After the key is assigned to a particular label, users can press that key whenever the mouse is over a photo. At that point, the spot that the cursor was over will be annotated with the specified label. When the expert user has a good idea about who or what is to be annotated, then the hotkeys can be put to work very efficiently. Instead of having to find the name desired and drag the name over to the position, the user can simply position the mouse and press the hotkey. This method is especially useful when a photo collection has only a few names that appear frequently.

Popup menu annotation (Figure 2(c)) also aims at reducing mouse movement in selecting labels to be used for annotation. This method offers a menu when the right mouse button is pressed on a picture. The menu consists of the currently selected labels in the list so

that users can annotate pictures with a label in the menu at the position the right mouse button was pressed. This mechanism was revised again later as “Label Paint Annotation” so that a photo can be annotated with the currently selected labels whenever users click on a photo without selecting a popup menu.

In addition to the proposed methods for improving the speed in annotating individual photos, we also designed two additional methods for improving the efficiency of bulk annotation as follows:

- Checkbox Annotation: Select one or more photos and click the check box next to the label in the list (Figure 1).
- Drag-drop Annotation: Select one or more photos and drag and drop the selected labels onto the photo. All the photos that were selected get annotated with the labels.

Another user study [Jung, 2000] was conducted to examine the differences in annotation performance and users’ satisfaction among the enhanced direct annotation methods. The results suggest that while there are some slight differences among the direct annotation methods, for the most part there is not enough of a significant distinguishing factor to proclaim one as the most efficient or the most rewarding. However, a further study of expert users may be necessary to verify this and to determine if there are conditions under which some approaches are more valuable.

If the methods aren’t significant, then the best option may be to include as many methods as possible while allowing a variety of options to let the users customize the methods.

However it raises the question of how much the user can learn at first; if presented by too many options in the beginning, the user may become confused and/or frustrated.

Therefore it may be optimal to use the multi-layered approach [Kang et al, 2003] when designing the initial interface.



(a) Split menu annotation



(b) Hotkey annotation



(c) Popup menu annotation

Figure 2 Enhanced Direct Annotation Approaches within PhotoFinder

2.3 Semi-Automatic Annotation

The performance of direct annotation may be improved by supporting bulk annotation as described in the previous section. However this approach introduces two new challenges. First, unless photos to be annotated with the same labels are clustered together in some way so that they can be selected together, users need to manually select photos one by one before applying bulk annotation. Second, since multiple photos are annotated with a single interaction, the position of the labels in each photo cannot be explicitly specified by users.

To cope with these challenges, we explored a semi-automatic annotation strategy that takes advantage of human and computer strengths [Suh and Bederson, 2007]. The semi-automatic approach enables users to efficiently update automatically obtained metadata interactively and incrementally. Even though automatically identified metadata are compromised with inaccurate recognition errors, the process of correcting inaccurate information can be faster and easier than manually adding new metadata from scratch. In order to facilitate efficient bulk annotation, two clustering algorithms were introduced to generate meaningful photo clusters [Suh and Bederson, 2007]: *Hierarchical event clustering* and *Clothing based person recognition*. The first method clusters photos based on their timestamps so that each group of photos corresponds to a meaningful event. The hierarchical event clustering identifies multiple event levels in personal photo collection and allows users to choose the right event granularity. The second method uses a clothing-based human model to group similar-looking people together. The clothing-

based human model is based on the assumption that people who wear similar clothing and appear in photos taken within relatively short periods of time are very likely to be the same person.

To explore our semi-automatic strategies, we designed and implemented a prototype called SAPHARI (Semi-Automatic PHoto Annotation and Recognition Interface). The prototype provides an annotation framework focusing on making bulk annotations on automatically identified photo groups. The two automatic clustering algorithms provide meaningful clusters to users, and make annotation more usable than relying solely on vision techniques such as face recognition or human identification in photos. In addition, the clothing based person recognition automatically detects the positions of people in a photo, which can be used to position annotations. Interestingly, while we found that absolute performance using this approach wasn't much better than manual annotation, the user study showed that the semi automatic annotation was significantly preferred over manual annotation [Suh and Bederson, 2007]. And as computer vision techniques improve, the relative advantage of semi-automated techniques can only improve.

Figure 3 shows a prototype of SAPHARI which shows the clusters of identified people whose upper bodies are cropped from photos of the same event. Since the automatic clustering techniques can also have errors, we designed the GUI so that users can correct errors manually by dragging a photo into the correct group. After error correction, users can annotate multiple photos at once (also with position information) by dragging a label onto the group.

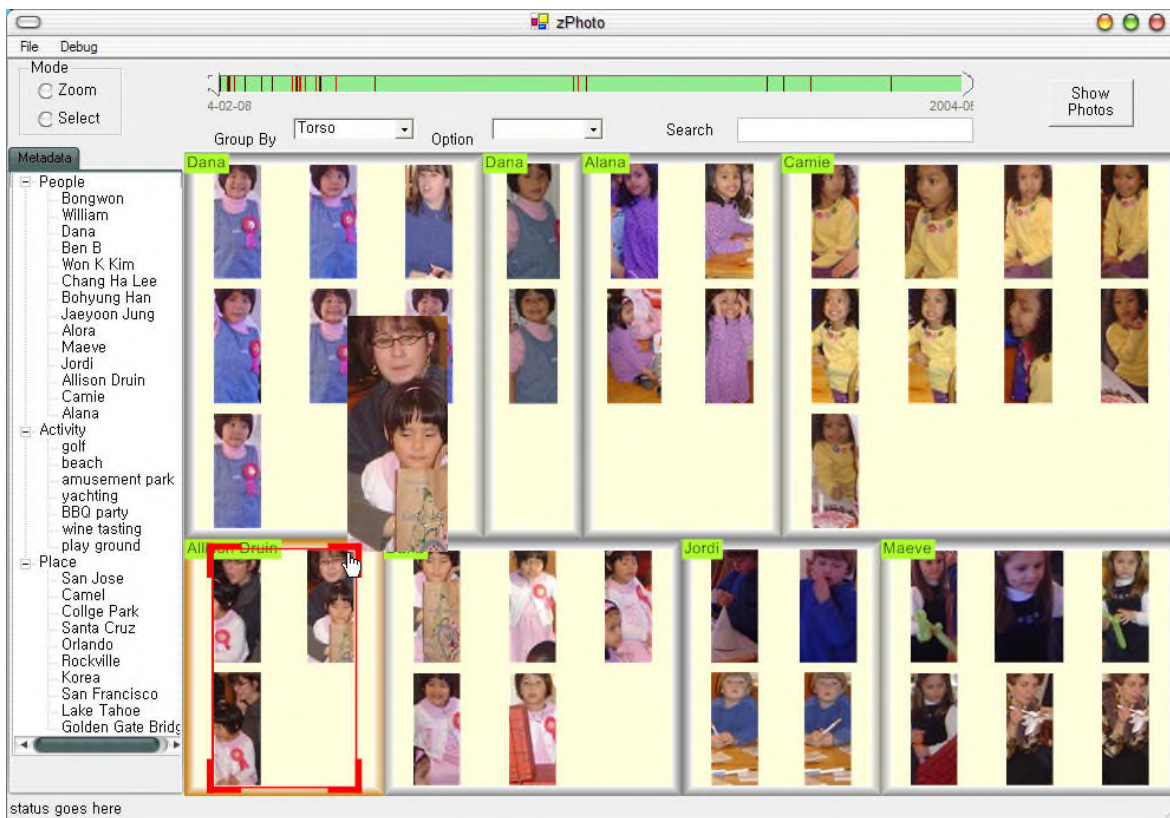


Figure 3: Identified people who are cropped from photos are laid out on the screen in the SAPHARI prototype. Similar looking people are clustered based on their clothing.

2.4 Image Annotation Discussion

There has been significant research on how to acquire useful metadata for images.

Many researchers have focused on automatic extraction of metadata by understanding images. For example, researchers has focused on automatic object detection such as face recognition, and content-based categorization and retrieval [Yang et al. 2002][Yoshitaka and Ichikawa 1999][Zhao et al. 2003]. However, such automatic techniques have achieved limited success so far when applied to personal photo management [Phillips et al. 2003] [Suh and Bederson 2007].

Rather than pure image-based approaches, some researchers used related context information to identify relevant metadata. For example, Naaman et al. [2005] used time and location information to generate label suggestions to identify the people in photos. Lieberman and Liu [2002] used relationships between text and photographs to semi-automatically annotate pictures.

On the other hand, web-based collaborative approaches to collecting metadata become popular. Web pages, online photographs, and Web links have been actively annotated with rich metadata through *tagging* (e.g. Flickr and Del.icio.us). Tagging is really just a simple kind of annotation where keywords can be associated with objects such as photos. However, the simple underlying approach belies the richness and power gained by broad communal access to these tags which enable cross-collection aggregation and searching along with innovative interfaces for creating and presenting those tags (i.e., tag clouds). With Flickr, a user can share personal photos with his/her friends, family and colleagues allowing the invited users to make annotations on shared photos. When users select photos and make them available to others, they seem to be willing to invest more effort in annotation [Shneiderman et al., 2006]. Also, by making them public, they invite others to comment and add annotations. We believe that this “folksonomy” based collaborative annotation has great potential for creating more useful metadata.

On the other hand, folksonomy based tagging system has a set of limitations. Folksomic tagging can be inconsistent because of its vocabulary problem [Furnas et al. 1987]. In folksonomy based systems, there is no bona fide standard for selecting a tag and users can choose to use any word as a tag. For example, one can use a tag, TV, while others

choose to use a tag, television. Furthermore, Furnas et al. [1987] showed that it is not always easy for one to come up with good descriptive keywords that can be shared. Tags, therefore, can be easily idiosyncratic and often causes meta-noise, which decreases the retrieval efficiency of systems.

Going further with collaborative annotation approaches, Von Ahn addressed the challenge by adding gaming to the mix. ESP Game [Von Ahn et al., 2006] combines a leisure game with practical goals. It lets people play an image guessing game that gathers image metadata as a side effect of them playing.

Another ongoing challenge for metadata management is where to store the metadata. It is crucial that users own their metadata just as much as they own their photos. Yet some companies (such as Apple with their iPhoto software) create a custom database that separates metadata from the photos leaving no easy way to get the annotations back for sharing or for use in other programs.

Software developers tend to like centralized metadata stores because they make it easier to write software that perform fast searches (and because they “lock in” customers to their products). However this approach is not necessary. Instead, we suggest storing all annotations and metadata using the “IPTC” standard format⁹ in the “EXIF” header [JEITA, 2002] of JPEG images. Then the application can create a cached copy of the metadata for efficient operations while leaving the “original” metadata with the photo. This is how we implemented PhotoMesa, and appears to be the same approach taken by

⁹ International Press Telecommunications Council <http://www.iptc.org>

some other commercial software such as Picasa and ACDSee¹⁰. However, even this approach has some challenges due to the following limitations:

- Not every image format supports metadata fields in the file headers.
- Metadata standards (i.e., EXIF/IPTC) do not have rich enough fields for objects such as people and hierarchical categories.
- Some operating systems make it difficult and possibly dangerous to modify the image header without changing the file modification date (which users sometimes use to determine when a file was created).

While the issues and challenges relating to metadata standards are significant, they are beyond the scope of this article. We thus raised these few issues to point out the importance of getting it right since it has such a dramatic affect on user experience. We therefore call on the industry to collaborate more openly in developing consistent and rich photo metadata standards addressing at least the above issues.

Technological change in society is rapid. There are new opportunities that we do not explore in this article. For example, the growing number of cell phones and even dedicated cameras that offer voice recording and network connectivity offer new possibilities such as supporting annotation at the moment the photo is captured. On the other hand, collaborative tagging has shown users are willing to make annotations when certain conditions are met. Further research is needed to investigate how the approaches we discuss in this article apply to those new settings.

¹⁰ ACD Systems <http://www.acdsee.com>

Even though the importance of photo metadata is well known, the development of systems to support annotation has not been so successful. That is partly because there have been many technical difficulties in automatic content analysis approaches, and partly because software designers have not always understood that many users are in fact willing to spend a substantial amount of energy annotating photos. In this section, we have demonstrated a few examples that show how novel interface designs based on a careful understanding of users can facilitate photo annotation and encourage users to annotate photos personally or collaboratively. In addition, we have explored issues that should be considered and resolved in order to encourage more researchers and developers to put their efforts in designing innovative GUIs to facilitate the annotation process for users.

3. GUIs for Using Metadata in Image Management and Exploration

Once richer image metadata is available either by automatic image content analysis or by manual annotation, users can make use of it for various image management tasks including not only search, but also organization, meaning extraction, navigation, and distribution.

From our research experience with personal media management systems [Bederson, 2001; Bederson et al., 2002; Kang and Shneiderman, 2000, 2002, 2006; Khella and Bederson, 2004; Kules et al., 2004; Shneiderman and Kang, 2000; Shneiderman et al., 2002; Shneiderman et al., 2006; Suh and Bederson, 2007], we learned that managing personal media objects is a challenge for most users, who may struggle to understand, interpret, arrange, and use them. They wrestle with at least two major problems. First, most tools were designed and developed based on rigid and system-driven organizing metaphors (such as file-folder hierarchy). Second, those tools were not suitably designed to use metadata in exploring the media objects except search.

To address these challenges, we have designed a number of novel GUIs that can improve task performance as well as user satisfaction in exploring and managing personal media data. In this section, we explain how these approaches can help image management tasks and how we tried to deal with those challenges.

3.1 Organization

With rich image metadata, even if users can always find the images they need, they still frequently want to organize them for other reasons such as supporting future serendipitous browsing and to provide the satisfaction of putting their images in order [Balabanovic et al., 2002; Kuchinsky et al., 1999; Rodden and Wood, 2003]. Hence, image organization is one of the most common and important image management tasks for users.

One of the main challenges in designing a novel user interface for organizing images is how to let end-users represent and flexibly apply their conceptual models to their image collections. Users typically understand their data by constructing conceptual models in their minds. There is no unique or right model. Rather, the models are personal, have meaning for the individual who creates them, and are tied to specific tasks. Even in a simple personal photo library, images can be organized by timelines, locations, events, people, or other attributes, depending on users' conceptual models and specific tasks. Despite the diversity of users' conceptual models, the means available for users to organize and customize their information spaces are typically poor and driven mostly by storage and distribution models, not by users' needs.

Ben Shneiderman and the first author of this article tried to resolve this challenge by providing users an environment to customize their information space appropriately for their conceptual models and specific tasks. We introduced a model called *Semantic*

Regions [Kang and Shneiderman 2006] (Figure 4), which are query regions drawn directly on a two-dimensional information space. Users can specify the shapes, sizes, and positions of the regions and thus form the layout of the regions meaningful to them.

Semantic Regions are spatially positioned and grouped on the 2D space based on personally defined clusters or well known display representations such as a map, tree, timeline, or organization chart. Once the regions are created and arranged, users can specify semantics for each region using image metadata. The specified image metadata are conjunctively joined to form the semantics of a region.

In Figure 4, each region represents a person, and the regions are grouped into 5 clusters to represent different friend groups. When the images are dragged onto the constructed model, they are automatically placed in the appropriate regions based on the annotations. This metaphor is called *fling-and-flock*; that is, users fling the objects and the objects flock to the regions. If photos do not satisfy any of the semantics of regions, they are collected in the *remaining items region* located at the top of the panel.

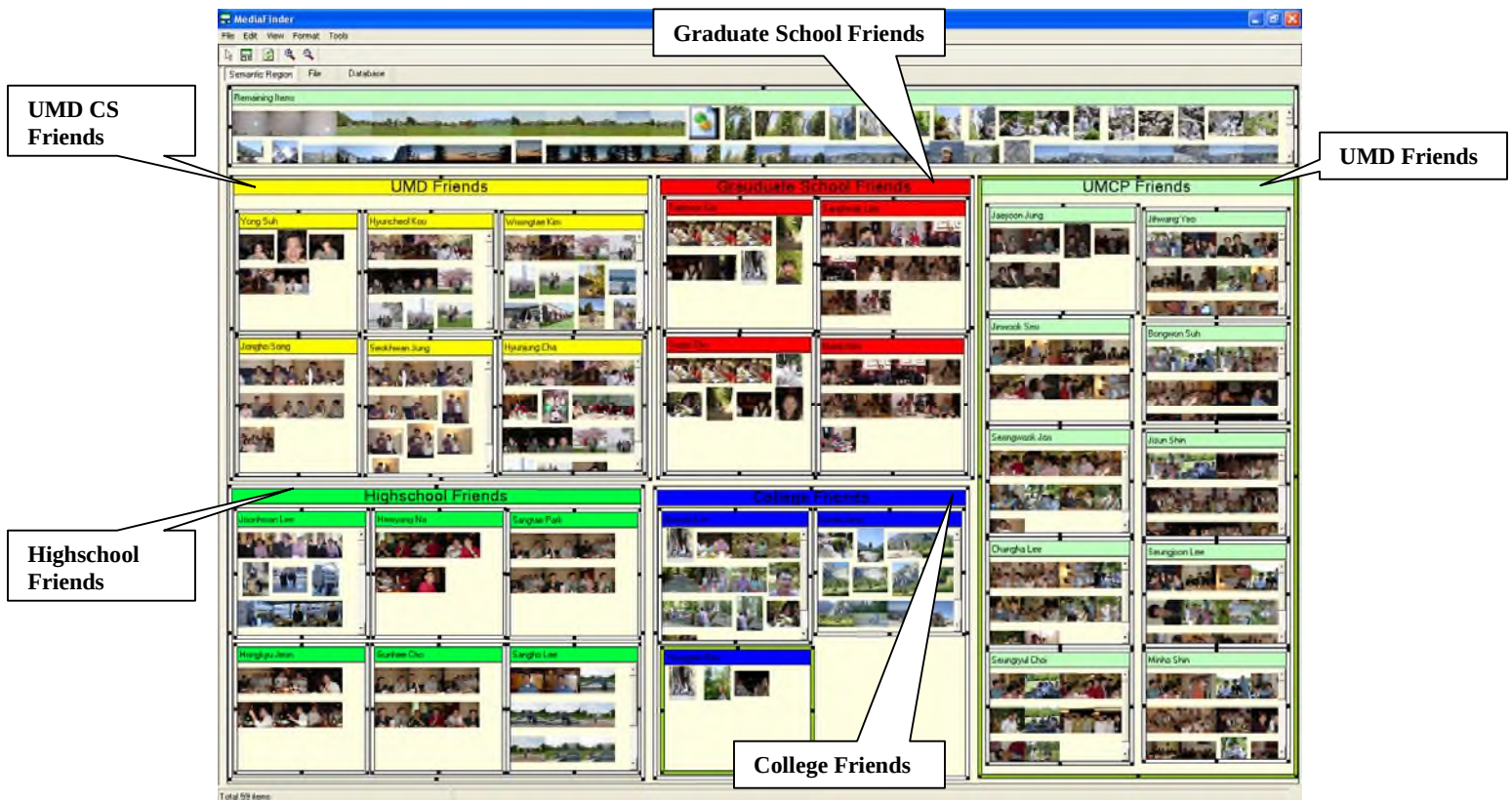


Figure 4: A friend group conceptual model: Each region represents a person and contains all the photos annotated with the name defined in it. The regions are grouped into 5 clusters to represent different friend groups (UMD friends -University of Maryland computer science students-, high school friends, graduate school friends, college friends, and UMCP friends-University of Maryland non-computer science students-). Each group has its own color to represent the different group of friends.

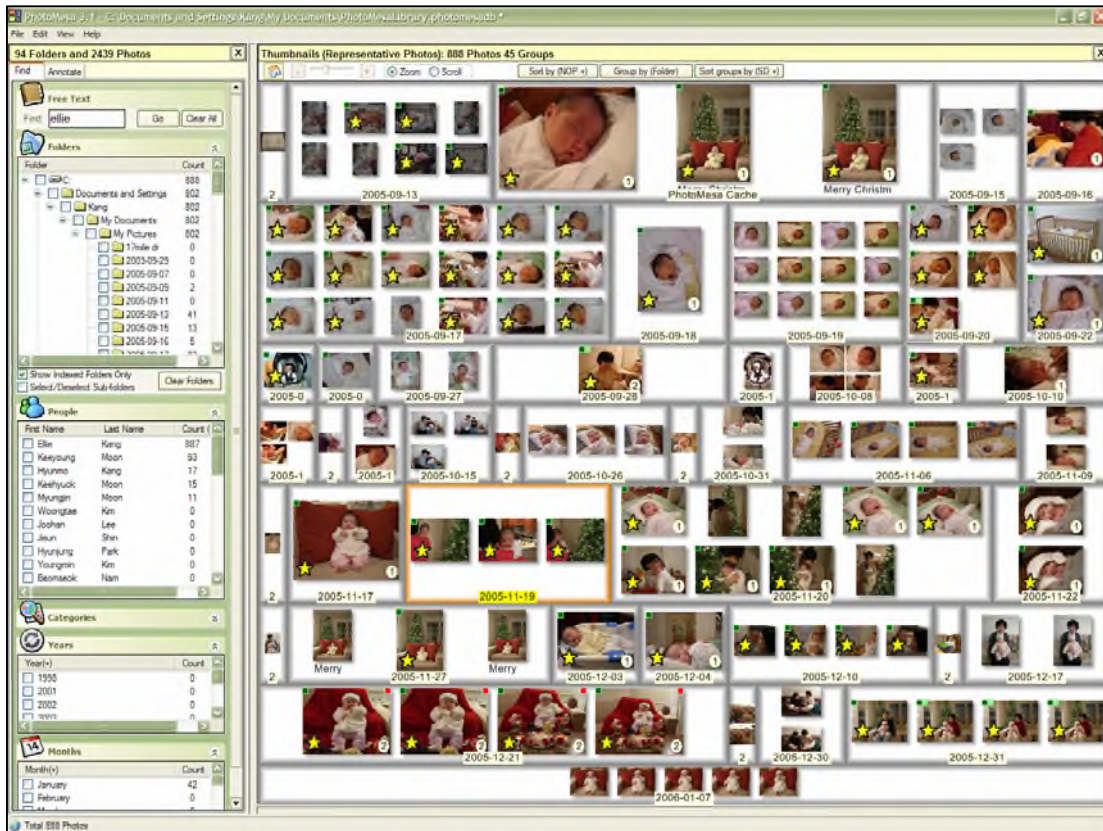
Small usability studies of the *Semantic Region* approach were conducted to discover potential user interface improvements, and to gain a deeper understanding about users' ability to understand, construct, and use semantic regions. The studies mainly focused on qualitative attributes such as user satisfaction rather than task performance. From these studies [Kang and Shneiderman, 2003], we learned that users liked the idea of spatial organization and dynamic grouping but they were not too interested in constructing their own models because it took too much time and effort. Users wanted rich built-in templates rather than user-defined templates, or more automatic features to reduce the steps for region construction and semantic specification. Based on these studies, we

simplified this approach by removing steps from the region construction and semantic specification while supporting automatic grouping and layout based on image metadata.

In a different project, we created an alternative and simplified approach for displaying groups of photos. This tool, called PhotoMesa, was developed as a successor to PhotoFinder by adding a zoomable user interface and the ability to show multiple groups of photos on the screen at the same time [Bederson, 2001; Bederson et al., 2002]. In PhotoMesa, users can dynamically control how photos are grouped based on available metadata (**Figure 5(a)**). For example, photos can be grouped by the folder they belong to, by when the photos were taken (month or year), by people in the photo, or by any user-defined categories (locations, events, objects, etc.) that the photos are annotated with. The groups are automatically placed using quantum treemap space-filling algorithm [Bederson et al., 2002], and the groups can also be sorted by photo metadata within a group such as number of photos in each group, starting date of photos in the group, title of the group, and so on. Each photo can be contained in multiple groups as long as it satisfies the grouping category.

Another approach created by Ben Shneiderman and Jack Kustanowitz is designed for groups of photos that have some relationship among them. In this case, it can be advantageous to show this relationship in the layout, rather than only in textual captions. **Figure 5(b)** shows a novel bi-level hierarchical layout, in which one region is designated for primary content, which can be a single photo, text, graphic, or combination [Kustanowitz and Shneiderman, 2005, 2006]. Adjacent to that primary region, groups of

photos are placed radially in an ordered fashion, such that the relationship of the single primary region to its many secondary regions is immediately apparent. A compelling aspect is the interactive experience in which the layout is dynamically resizable, allowing users to rapidly, incrementally, and reversibly alter the dimensions and content.



(a) PhotoMesa interface enables users to dynamically group photos based on photo metadata. The group layout is automatically generated using space filling algorithm but they can also be sorted by the metadata of group photos. PhotoMesa enables users to control the visible photos in various ways. In this example, PhotoMesa shows only the representative photos of each group.



(b) Bi-level hierarchical layout shows the photo groups taken in 6 locations during a trip by placing the groups clockwise in chronological order(left). A real estate browser with 5 communities and the price ranges of selected homes is shown on the right.

Figure 5: Organization of Personal Photo Collections

3.2 Navigation and Browsing

Image browsing is important for a number of reasons. First of all, no matter what information retrieval system is being used, the user has to browse the results of the search. It is certainly important to build query systems that help users get results that are as close to what is wanted as possible. However, there will always be images that need to be browsed visually to make the final pick. The second reason is that sometimes people browse images just for the pleasure of looking at those images, and they often do it with other people. This is especially true for personal photos. Looking at home photos has a lot of overlap with traditional retrieval systems. People still want to be able to find photos of particular people and events, etc. However, they are less likely to be time pressured to find a particular photo, and more likely to be interested in serendipity (that is, finding photos they weren't looking for) [Balabanovic et al., 2000; Bederson, 2002; Rodden and Wood, 2003, Shneiderman et al., 2006].

We also designed PhotoMesa to support this serendipitous style of use. It uses a zoomable user interface (Figure 6) to enable users to see all the images in a single view and concentrate on the images while browsing photos by zooming in or out without the need for interacting with other UI widgets such as scroll bar or popup window. This design is also based on the characteristics of personal image collections. As users have seen them at least once, they already have rough idea about their images and metadata. Hence, they often want to start browsing from the overview of their whole collection without worrying about missing some photos by accident (because they scrolled outside the currently visible region).

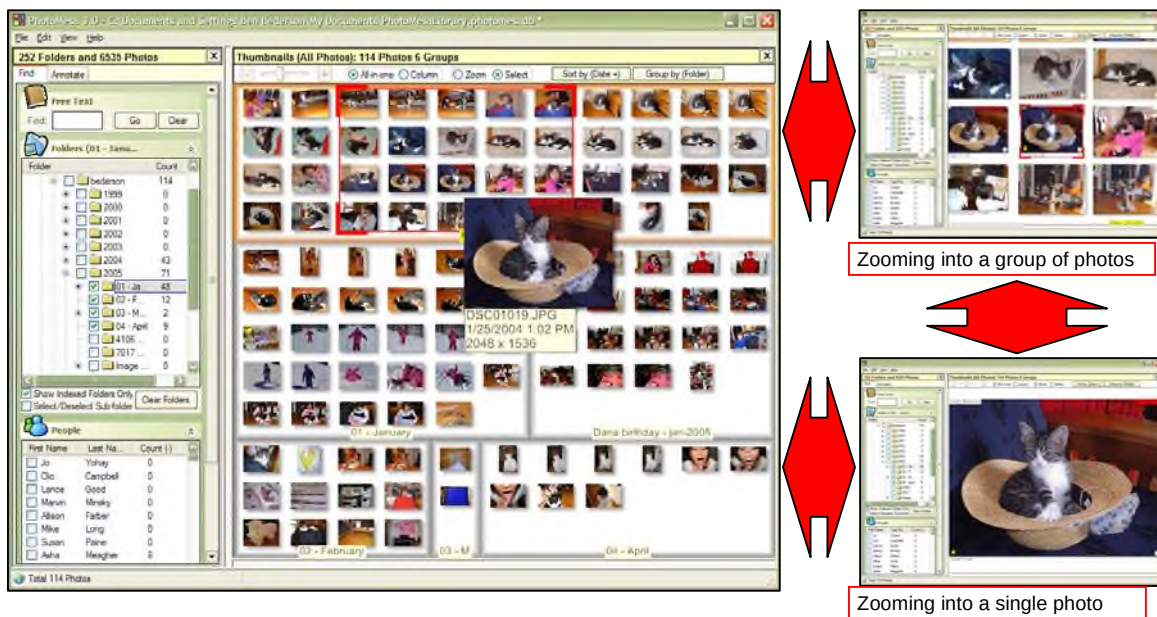


Figure 6: PhotoMesa shows all photos in a single view. The layout of photos depends on how many groups and photos have been loaded. A larger preview of any photo can be obtained by moving the mouse over one of the small thumbnails. PhotoMesa uses a Zoomable User Interface (ZUI) and supports browsing of photos by zooming in or out. The order as well as the group that photos are displayed in can be controlled dynamically by photo metadata.

This zoomable browsing approach was further enhanced in terms of system performance and users' cognitive workload by supporting semantic zooming [Bederson et al., 1996]. In the PhotoMesa interface, if there are too many displayed for users to recognize at once, PhotoMesa has the option of displaying just representative photos (Figure 4a). A layout such as *Time quilt* [Huynh et al., 2005] (Figure 8(a)) showed the improvement in photo browsing performance even more by making use of linear timeline layout on top of the 2D space-filling techniques with semantic zooming. This layout is designed to convey a temporal order while making reasonable use of screen space by packing a timeline layout into a rectangular screen space using a "line break" algorithm. Their user study shows that this combined layout enabled users to complete some photo browsing tasks faster than the quantum treemap layout and the timeline layout.

On the other hand, a new tool called PhotoSynth [Snavely et al, 2006] (Figure 8(b)) uses 3D space for photo browsing. The system takes a large collection of photos of a place or an object, analyzes them for similarities, and displays them in a reconstructed three-dimensional space so that users can interactively browse and explore large collections of photos of a scene using a 3D interface.

In addition to the design of novel browsing interfaces, we also developed a novel thumbnail generation algorithm which aims to reduce users' cognitive workload during browsing [Suh et al., 2003]. Using thumbnails has a tradeoff; bigger thumbnails use more screen space while smaller thumbnails are hard to recognize. Automatic thumbnail cropping is a method to find the optimal part of an image for thumbnail generation. The algorithm finds the informative portion of images based on a saliency model and cuts out the non-salient part of images. Thumbnail images generated from the cropped part of images increases users' recognition and helps users in visual search. On the other hand, the face detection based cropping technique [Suh et al., 2003] shows how semantic information can be used to enhance thumbnail generation further (when that semantic information about an image is known). For example, using face detection can more effectively create thumbnails with the images of people, which was demonstrated to significantly increase users' recognition and visual search performance.

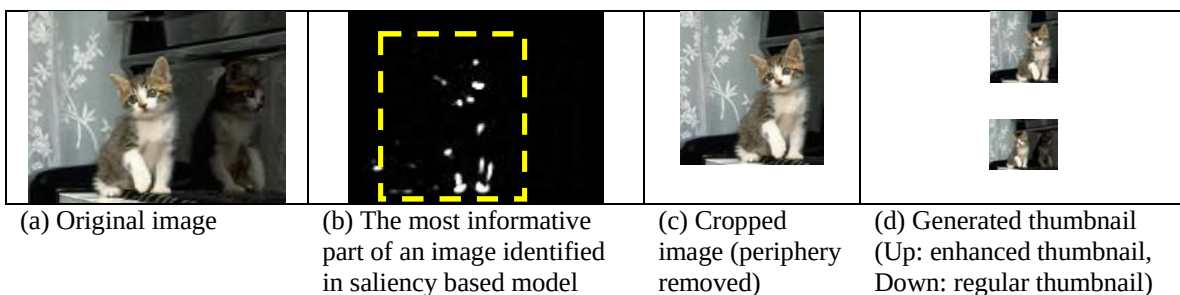
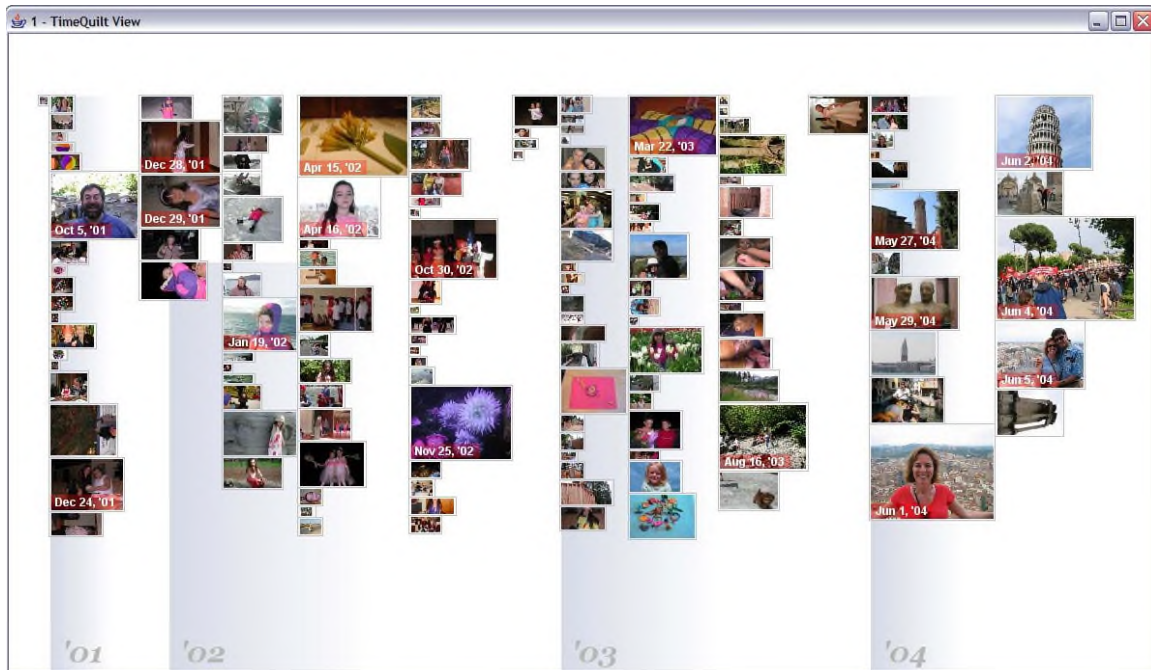


Figure 7 The saliency based model identifies the most informative part of an image. Thumbnail images created out of the core part help users' recognition and visual search performance.



(a) Time quilt – a layout designed to convey temporal order while making better use of screenspace than a timeline, showing approximately 5,500 photos with representative thumbnail overview



(b) PhotoSynth takes a large collection of photos of a place or an object, analyzes them for similarities, and displays them in a reconstructed three-dimensional space.

Figure 8 Photo Browsing Interfaces

3.3 Search and Meaning Extraction

There is no doubt that search is a hugely important photo management task. Users may be looking for photos of their grandmother, their trip to Disney World, or a friend's wedding. Interactive visualization to select and view dozens or hundreds of photos extracted from tens of thousands has become a popular strategy [Shneiderman et al., 2006]. In addition to simple photo search based on annotation, users sometimes want to extract meaning from their photo collections such as "Who else appeared in the pictures of my grandmother and me over 20 years and when and where did I meet her?", "Who appeared both in my friend's wedding pictures and in my wife's friend's wedding pictures?", "How many photos did I take for my daughter each year?, Were there any special events since she was born? Were there any seasonal differences in photo distribution?", and so on. This kind of meaning extraction from personal photo collections becomes more common and popular tasks as users have more and more digital photos and richer metadata. However, little work has been done so far in this area.

One effort in this area, though, was done in the *Semantic Region* project [Kang and Shneiderman, 2006] described earlier in this section provides an interactive visualization technique called *region brushing* to support meaning extraction task by visualizing the relationships among the semantic regions. Region brushing is used to highlight the personal media items contained in multiple regions simultaneously (Figure 9(a)) and is often used for depicting both the intra-relationships and the inter-relationships among models. Since a personal media item can be contained in multiple semantic regions, the existence as well as the amount of the shared personal media items among the regions

well represents their relationships. In Figure 9(b), three models, HCI researcher, US map, and CHI conference calendar are combined together. Since a photo can be contained in multiple regions across the different models, many questions concerning the interrelationships of models can be answered through region brushing. Such questions might be “Find the name of the conferences that all nine people participated in and where was the conference held?”, “What was the name of the conference held in Atlanta, Georgia, and who did not appear in the photos taken at this conference?”, “Find the conferences in which Ben Shneiderman did not appear in the photos and where was the conference held?”, and so on. Furthermore, constructing new semantic regions recursively inside a region (e.g. location regions inside a person region) enables users to dynamically regroup the photos contained in a region and extract meanings.

Simple techniques such as *query preview* [Plaisant et al., 1999] or *faceted search* [Yee et al., 2003] can also help users extract meanings as well as filter down a collection and show potential targets from their personal collections. PhotoMesa applies these approaches by letting users search for photos using the Find tab (left panel in Figure 10) in six ways: with keywords; by the folder a photo is in; by the people in a photo; by the categories a photo was annotated with; and by the year or month a photo was taken. If a new photo set is searched by combining those six ways, each search panel previews the distribution of the searched photos so that users can see how many photos are from each folder, how many people (and who) appear in the searched photos, where the searched photos were taken (or how many photos were from other user-defined categories), when they were taken (month and year). With this feature, users can get information about the searched photos and use it for further exploratory search.

3.4 Sharing and Distribution

Photo sharing is becoming one of the most common activities of digital photo management. Our earlier work provided interesting insights about how individuals and communities share their personal photographs. For example, a client-server model of PhotoFinder, PhotoFinder Kiosk (which was installed at CHI 2001), enabled members of a community of HCI professionals to enjoy, share and annotate photos at an interactive public exhibit [Kules et al., 2004, Shneiderman et al., 2002]. On the other hand, as mentioned earlier, there is a tremendous commercial activity in this area. There are numerous popular photo sharing websites such as Flickr, Picasa Web Albums, etc.

We previously mentioned some of the trade-offs of desktop applications and web-based photo-sharing sites. However, the current bifurcation requires users to pick one approach or the other to manage their photos. We believe that this split is an artifact of business requirements and software development practices, and that with a little more creativity and effort, it is possible to build a hybrid solution that bridges both worlds providing the advantages of each. For example, Sharpcast¹¹ enables users to share, access, and organize their photos either through local file systems and web sites. We also have heard of people using independent file synchronization tools, such as FolderShare¹² to keep multiple copies of their photos synced with metadata between different machines including web servers. All this points to the necessity of a hybrid approach. From a user's perspective, it doesn't make any sense that they should have to choose in advance whether they should commit to storing photos locally or online.

¹¹ Sharpcast, Inc. <http://www.sharpcast.com>

¹² Microsoft Inc. <http://www.foldershare.com>

To assess the usefulness of the hybrid approach, we built a prototype extension to PhotoMesa that hides the distinction between local and web access from the user. Since photos almost always start locally, after being transferred from a camera to a computer, we assume users will open the photos regularly in their local tool. Then the tool continuously synchronizes the local photos with the Flickr website using its public API in the background. In this manner, all the user's photos are also available within their private Flickr account. Furthermore, any changes or annotations to the photos on Flickr are synced back to the desktop application the next time they are accessed locally. In addition, full searches of Flickr's remote database can be performed within PhotoMesa the same way local searches are made, and remote photos are shown grouped by whatever metadata Flickr makes available.

There is a largely unmet need of users to be freed from the shackles of tightly controlled applications that limit access to their photos and metadata. In the increasingly decentralized world of information, that iron-fisted approach is just not sustainable any more.

4. Conclusion

The world of digital photo management has come a long way in the past few years, and we are pleased to see that the basic needs of most personal photo users are now being met. However we aren't done yet. There remain large gaps in meeting users' fuller needs such as recording richer annotations that are more accurate and broadly available. And it is crucial to support seamless sharing of photos and their stories in a way that lets multiple parties participate without losing ownership of the photos and metadata.

It is not clear how well current approaches will scale up. With the ever-shrinking cost of storage at the same time that cameras are becoming ubiquitous, we can expect that over an individual's life, it will be common to manage tens or even hundreds of thousands of photos.

The challenge of applying automated or human-assisted analysis techniques to these photos in a way that remains understandable and usable is ongoing. It also remains a challenge to find the right balance between local and web access that enables people to have reliable long-term access to their full resolution photos in a way that doesn't lock them into any particular vendor's system. People must also be able to maintain their privacy and control over photos while sharing and collaboratively editing the photos and stories about them. And when we envision the addition of audio and video to these collections which is already happening, all of these challenges grow even further.

The transition to digital photography has already largely happened. This implies that consumers are satisfied enough with current solutions – or at least prefer digital to film. However that doesn't mean that their long-term problems are solved, or that things are good as they could and should be. It is our job as designers and developers of these systems to address these challenges.

In this article, we described a range of projects that attempt to improve some of the problems already described. By integrating better annotation tools, visual presentation and query systems, and through the use of more open metadata storage approaches, we have the opportunity to take the growing digital photo market and make it really explode – while providing better systems for all of us to use with our own dear personal photos. Some directions not covered in this article, such as integrating annotations into other life activities (such as blogging, diary writing, email, etc.) also offer promise. In the end, we expect a rich ecosystem of photo-related technologies will continue to evolve that support the diverse set of users that take photos.

5. Reference

1. Balabanovic, M., Chu, L.L., and Wolff, G.J., Story-telling With Digital Photographs, *In Proceeding of Human Factors in Computing Systems (CHI 2000)*, ACM Press, pp. 564-571, 2000.
2. Bederson, B.B., PhotoMesa: A Zoomable Image Browser using Quantum Treemaps and Bubblemaps, *In Proceeding of ACM Conference on User Interface and Software Technology (UIST 2001)*, ACM Press, pp. 71-80, 2001.
3. Bederson, B.B., Shneiderman, B., and Wattenberg, M., Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies, *ACM Transactions on Graphics*, 21, 4, pp. 833-854. *
4. Chalfen, R., Snapshot Versions of Life, Bowling Green State University Popular Press, Ohio, 1987.
5. Chellappa, R., Wilson, C.L., and Sirohey, S., Human and Machine Recognition of Faces: A Survey, *In Proceeding of the IEEE*, Vol. 83, pp. 705-740, 1995.
6. Furnas, G.W., Landauer, T.K., Gomez, L.M., and Dumais, S.T., The Vocabulary Problem in Human-System Communication, *Communication of ACM*, 30, pp. 964-971, 1987.
7. Goto, Y., Jung, J., Ma, K., and McCaslin, O., The Effect of Direct Annotation on Speed and Satisfaction, SHORE, UMD, 2000.
<http://www.otal.umd.edu/SHORE2000/annotation/>

8. Huynh, D., Drucker, S., Baudisch, P., Wong, C., Time Quilt: Scaling up Zoomable Photo Browsers for Large, Unstructured Photo Collections. *In Extended Abstracts of Human Factors in Computing Systems (CHI 2005)* , Portland, OR, pp. 1937-1940, 2005.
9. JEITA (Japan electronic and information technology industries association), Exchangeable image file format for digital still cameras: Exif Version 2.2, 2002.
<http://www.exif.org>
10. Jung, J., Empirical Comparison of Four Accelerators for Direct Annotation of Photos, 2000. <http://www.cs.umd.edu/hcil/photolib/paper/cmssc498paper.doc>
11. Kang, H. and Shneiderman, B., Dynamic Layout Management in a Multimedia Bulletin Board, *In Proceeding of IEEE International Symposium on Human-Centric Computing Languages and Environments (HCC 2002)*, 2002. *
12. Kang, H., Plaisant, C., and Shneiderman B., Helping Users Get Started with Visual Interfaces: Multi-Layered Interfaces, Integrated Guidance and Video Demonstrations *In Proceedings of Digital Government Conference (DG2003)*, 2003. *
13. Kang, H., Shneiderman, B., Exploring Personal Media: A Spatial Interface Supporting User-Defined Semantic Regions, *Journal of Visual Language and Computing*, volume 17, issue 3, pp. 254-283, 2006. *
14. Kang, H., Shneiderman, B., Visualization Methods for Personal Photo Collections: Browsing and Searching in the PhotoFinder, *In Proceeding of IEEE International Conference on Multimedia and Expo (ICME2000)*, 2000. *

15. Khella, A., Bederson, B.B., Pocket PhotoMesa: A Zoomable Image Browser for PDAs., *In Proceeding of Mobile and Ubiquitous Multimedia (MUM 2004)*, ACM Press, pp. 19-24, 2004.
16. Kuchinsky, A., Pering, C., Creech, M.L., Freeze, D., Serra, B., Gwizdka, J., FotoFile: A Consumer Multimedia Organization and Retrieval System, *In Proceeding of ACM Conference on Human Factors in Computing Systems (CHI99)*, pp. 496-503, 1999.
17. Kules, B., Kang, H., Plaisant, P., Rose, A., Shneiderman, B., Immediate Usability: A Case Study of Public Access Design for a Community Photo Library, *Journal of Interacting With Computers*, volume 16, num 6, pp. 1171-1193, Elsevier, 2004. *
18. Kustanowitz, J. and Shneiderman, B., Meaningful presentations of photo libraries: Rationale and applications of bi-level radial quantum layouts, *In Proceeding of Joint Conference on Digital Libraries*, ACM Press, 2005. *
19. Kustanowitz, J., Shneiderman, B., Hierarchical layouts for photo libraries, *IEEE MultiMedia*, 13, 4, pp. 62-72, 2006. *
20. Lieberman, H. and Liu, H., Adaptive linking between text and photos using common sense reasoning, *In Proceeding of Adaptive Hypermedia and Adaptive Web Systems*, Malaga, Spain, 2002.
21. Naaman, M., Yeh, R. B., Garcia-Molina, H., Paepcke, A., Leveraging Context to Resolve Identity in Photo Albums, *In Proceeding of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005)*, 2005.
22. Phillips, P. J., Grother, P. J., Michaels, R. J., Blackburn, D. M., Tabassi, E., Bone, J. M., Face recognition vendor test 2002: Evaluation report., NISTIR 6965, 2003.

23. Plaisant, C., Venkatraman, M., Ngamkajornwiwat, K., Barth, R., Harberts, B., Feng, W., Refining query previews techniques for data with multivalued attributes: The case of NASA EOSDIS, *In Proceeding of Advanced Digital Libraries 99 (ADL'99)*, IEEE Computer Society Press, pp. 50-59, 1999.
24. Rodden, K. and Wood, K., How do People Manage Their Digital Photographs?, *In Proceeding of ACM Conference on Human Factors in Computing Systems (CHI 2003)*, Fort Lauderdale, 2003.
25. Shneiderman, B., Bederson, B.B., and Drucker, S.M., Find that photo!: interface strategies to annotate, browse, and share., *Communications Of the ACM*, 49, 4, pp. 69-71, 2006. *
26. Shneiderman, B., Designing the User Interface: Strategies for Effective Human-Computer Interaction, 4th Edition, *Addison Wesley Longman*, Reading, MA, 2005. *
27. Shneiderman, B., Direct manipulation: a step beyond programming languages, *IEEE Computer* 16(8), pp. 57-69, 1983. *
28. Shneiderman, B., Kang, H., Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos, *In Proceeding of International Conference Information Visualization (IV2000)*. IEEE Computer Society, pp. 88-95, 2000. *
29. Shneiderman, B., Kang, H., Kules, B., Plaisant, C., Rose, A., and Ruchair, R., A photo history of SIGCHI: evolution of design from personal to public, *ACM Interactions*, Volume 9, Issue 3, pp. 17-23, 2002. *
30. Snavely, N., Seitz, S. M., Szeliski, R., Photo tourism: exploring photo collections in 3D, *ACM Transactions on Graphics (TOG)*, Volume 25 Issue 3, pp. 835-846, 2006.

31. Suh, B., Bederson, B.B., Semi-Automatic Image Annotation Strategies Using Event Based Clustering and Clothing Based Person Recognition, *Journal of Interacting with Computers*, Elsevier, 2007.
32. Suh, B., Ling, H., Bederson, B.B., Jacobs, D., Automatic Thumbnail Cropping and its Effectiveness, *In Proceeding of ACM Conference on User Interface and Software Technolgy* (UIST 2003), pp. 95-104, 2003
33. Von Ahn, L. and Dabbish, L., Labeling Images with a Computer Game, *In Proceeding of ACM Conference on Human Factors in Computing Systems* (CHI 2004), ACM Press, pp. 319-326, 2004.
34. White, R., Kules, B., Bederson, B.B., Exploratory Search Interfaces: Categorization, Clustering and Beyond, *SIGIR Forum*, volume 39, issue 2
35. Yang, M., Kriegman, D., and Ahuja, N., Detecting Faces in Images: A Survey, *IEEE Transactions on Pattern Analysis and Mach Intelligence*, 24(1), pp. 34-58, 2002.
36. Yee, K-P., Swearingen, K., Li, K., and Hearst, M., Faceted Metadata for Image Search and Browsing, *In Proceeding of ACM Conference on Human Factors in Computing Systems* (CHI 2003), ACM Press, 2003.
37. Yoshitaka, A., and Ichikawa, T., A Survey on Content-Based Retrieval for Multimedia Databases., *IEEE Trans on Knowledge and Data Engineering*, 11(1): pp. 81-93, 1999.
38. Zhao, W., Chellappa, R., Philips, P.J., Rosenfeld, A., Face Recognition: A Literature Survey, *ACM Computing Surveys*, Vol 35(4), pp. 399-458, 2003.