Direct Annotation:
A Drag-and-Drop Strategy for Labeling Photos

Ben Shneiderman, Hyunmo Kang
Dept. of Computer Science, Human-Computer Interaction Laboratory,
Institute for Advanced Computer Studies & Institute for Systems Research
University of Maryland, College Park, MD 20742 USA
{ben, kang}@cs.umd.edu

Abstract

Annotating photos is such a time-consuming, tedious and error-prone data entry task that it discourages most owners of personal photo libraries. By allowing users to drag labels such as personal names from a scrolling list and drop them on a photo, we believe we can make the task faster, easier and more appealing. Since the names are entered in a database, searching for all photos of a friend or family member is dramatically simplified. We describe the user interface design and the database schema to support direct annotation, as implemented in our PhotoFinder prototype.

Keywords: direct annotation, direct manipulation, graphical user interfaces, photo libraries, drag-and-drop, label placement

1. Introduction

Adding captions to photos is a time-consuming and error prone task for professional photographers, editors, librarians, curators, scholars, and amateur photographers. In many professional applications, photos are worthless unless they are accurately described by date, time, location, photographer, title, recognizable people, etc. Additional annotation may include details about the photo (for example, film type, print size, aperture, shutter speed, owner, copyright information) and its contents (keywords from controlled vocabularies, topics from a hierarchy, free text descriptions, etc.). For amateur photographers, annotations are rarely done, except for the occasional handwritten note on the back of a photo or an envelope containing a collection of photos.

For those who are serious about adding annotations, the common computer-based approach is to use database programs, such as Microsoft Access, that offer form fill-in or free text boxes and then store the information in a database. Data entry is typically done by typing, but selecting attribute values for some fields (for example, black&white or color film) is supported in many systems. Of course, simpler tools that provide free-form input, such as word processors, spreadsheets, and other tools are used in many situations. Captions and annotations are often displayed near a photo on screen displays, web pages, and printed versions. Software packages (Kodak PhotoEasy, MGI PhotoSuite, Aladdin Image AXS, etc.) and web sites (Kodak's photonet, Gatherround.com, shutterfly, etc.) offer modest facilities to typing in annotations and searching descriptions.

As photo library sizes increase the need and benefit of annotation and search capabilities grows. The need to rapidly locate photos of Bill Clinton meeting with Boris Yeltsin at a European summit held in 1998 is strong enough to justify substantial efforts in many news agencies. More difficult searches such as "agriculture in developing nations" are harder to satisfy, but many web and database search tools support such searches (Lycos, Corbis, etc.). Query-By-Image-Content from IBM, is one of many projects that uses automated techniques to analyze image (http://wwwqbic.almaden.ibm.com/). Computer vision techniques can be helpful in finding photos by color (sunsets are a typical example), identifying features (corporate logos or the Washington Monument), or textures (such as clouds or trees), but a blend of automated and manual techniques may be preferable. Face recognition research offers hope for automated annotation, but commercial progress is slow [1][2].

2. Related Work on Annotation

Annotation of photos is a variation on previously explored problems such as annotation on maps [3][4][5] in which the challenge is to place city, state, river, or lake labels close to the features. There is a long history of work on this problem, but new possibilities emerge because of the dynamics of the computer screen (Figure 1). However, annotation is usually seen as an authoring process conducted by specialists and users only chose

whether to show or hide annotations. Variations on annotation also come from the placement of labels on markers in information visualization tasks such as in tree structures, such in the hyperbolic tree [6] (Figure 2) or in medical histories, such as LifeLines [7] (Figure 3).

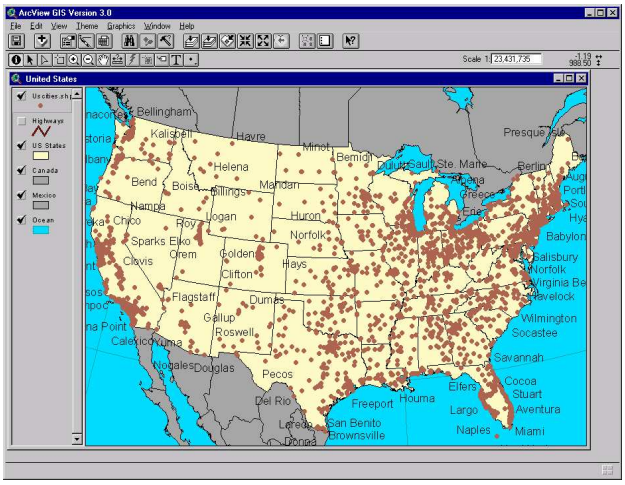

Figure 1. US Map with City Names

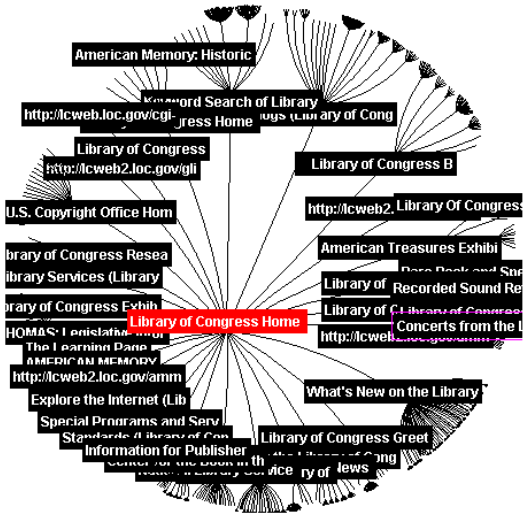

Figure 2. Hyperbolic Tree

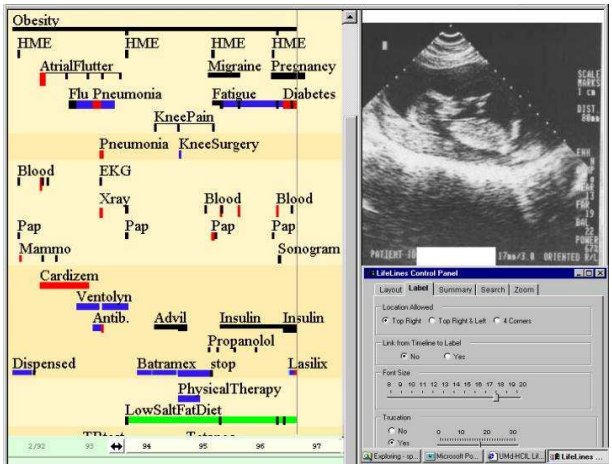

Figure 3. LifeLines Medical Patient History

Previous work on annotation focused on writing programs to make label placements that reduced overlaps [8], but there are many situations in which it is helpful for users to place labels manually, much like post-it notes, on documents, photos, maps, diagrams, webpages, etc. Annotation of paper and electronic documents by hand is also a much-studied topic with continuing innovations [9]. While many systems allow notes to be placed on a document or object, the demands of annotating personal photo libraries are worthy of special study [10]. We believe that personal photo libraries are a special case because users are concentrating on the photos (and may have a low interested in the underlying technology), are concerned about the social aspects of sharing photos, and are intermittent users. They seek enjoyment and have little patience for form filling or data entry.

3. The PhotoFinder Project

In the initial stages of our project on storage and retrieval from personal photo libraries (http://www.cs.umd.edu/hcil/photolib/), we emphasize collection management and annotation to support searching for people. This decision was based on our user needs assessment, reports from other researchers, and our personal experience that indicate that people often want to find photos of a friend or relative at some event that occurred recently or years ago [2][11]. Personal photo libraries may have from hundreds to tens of thousands of photos, and organization is, to be generous, haphazard. Photos are sometimes in neat albums, but more often put in a drawer or a shoebox. While recent photos are often on top, shuffling through the photos often leaves them disorganized. Some users will keep photos in the envelopes they got from the photo store, and more organized types will label and order them.

As digital cameras become widespread, users have had to improvise organization strategies using hierarchical directory structures, and typing in descriptive file and directory names to replace the automatically generated photo file numbers. Some software packages (PhotoSuite, PhotoEasy, etc.) enable users to organize photos into albums and create web pages with photos, but annotation is often impossible or made difficult. Web sites such as Kodak's Photo.net, Gatherround.com, etc. enable users to store collections of photos and have discussion groups about the collections, but annotation is limited to typing into a caption field. The pioneering effort of the FotoFile [2] offered an excellent prototype that inspired our work.

Our goal in the PhotoFinder project was to support personal photo library users. We developed a conceptual model of a library having a set of collections, with each collection having a set of photos. Photos can participate in multiple collections. Collections and individual photos can be annotated with free text fields plus date and location fields stored in a database (see Figure 6 for our Photo Library database schema). Our interface has three main windows:

- **Library viewer**: Shows a representative photo for each collection, with a stack representing the number of photos in each collection.
- **Collection viewer**: Shows thumbnails of all photos in the collection. Users can move the photos around, enlarge them all or individually, cluster them, or present them in a compact manner. A variety of thumbnail designs were prototyped and will be refined for inclusion in future versions.
- **Photo viewer**: Shows an individual photo in a resizable window. A group of photos can be selected in the Collection viewer and dragged to the Photo viewer to produce an animated slide show.

We also put a strong emphasis on recording and searching by the names of people in each photo. We believed that a personal photo library might contain repeated images of the same people at different events, and estimated 100-200 identifiable people in 10,000 photos. Furthermore we expected a highly skewed distribution with immediate family members and close friends appearing very frequently. The many-to-many relationship between photos and people is mediated by the Appearance relation (Figure 6) that stores the identification of all the people who appear in each photo.

Such a database would support accurate storage of information, but we recognized that the tedious data entry problem would prevent most users from typing in names for each photo. Furthermore, the inconsistency in names is quickly a problem with misspellings or variant names (for example, Bill, Billy, William) undermining the success of search.

A second challenge we faced was that the list of names of people appearing in a photo could often be 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.

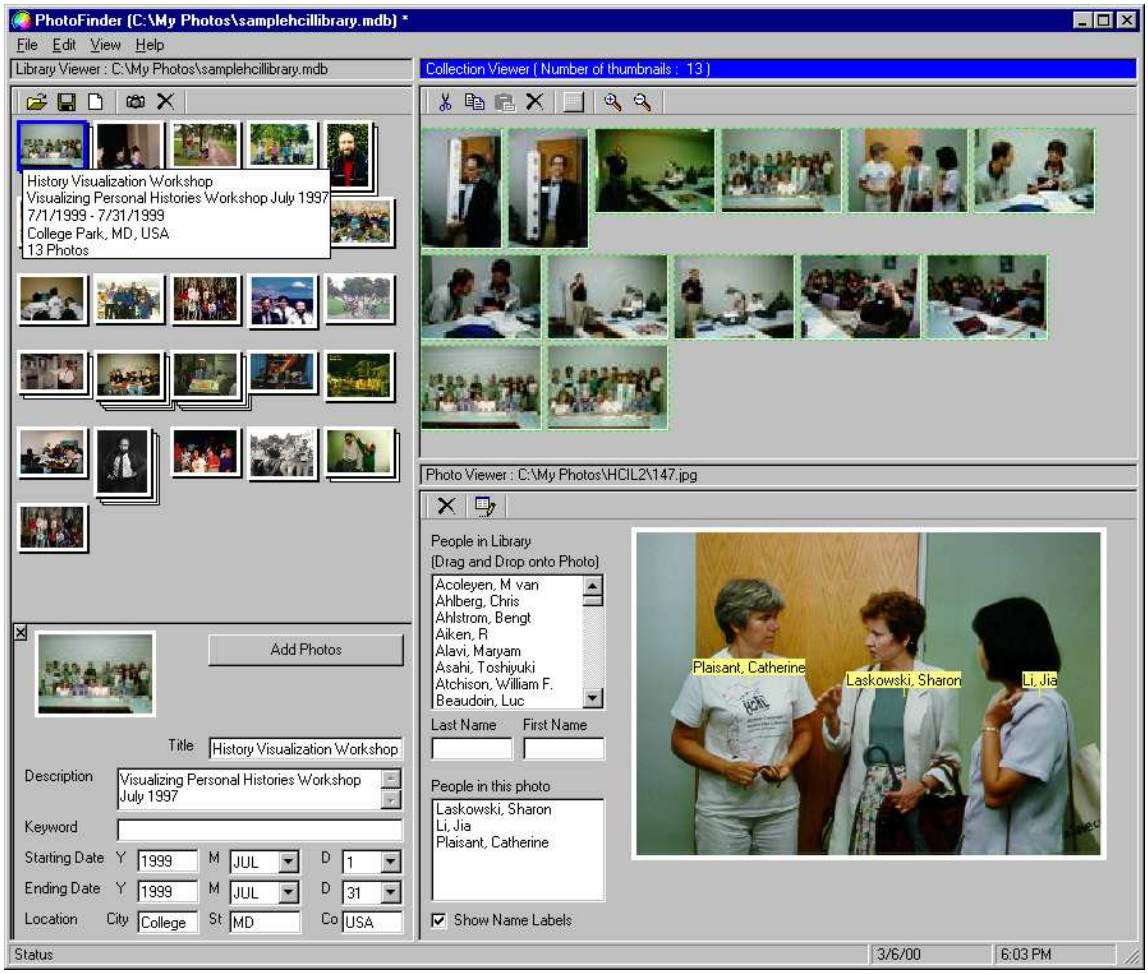

Figure 4. PhotoFinder1 display with Library Viewer on the left, Collection Viewer with thumbnails on the upper right, and Photo Viewer on the lower right.

4. Direct Annotation

To cope with these challenges we developed the concept of direct annotation: selectable, dragable labels that can be placed directly on the photo. Users can select from a scrolling or pop-up list and drag by mouse or touch screen. This applies direct manipulation principles [12] 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 in the Appearance relation with an X-Y location, based on an origin in the upper left hand corner of the photo.

This simple rapid process also allows users to annotate at will. 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 easy design and continuous annotation facility may encourage users to do more annotation. Figures 5 (a)-(f) show the process of annotation on a set of four people at a conference.

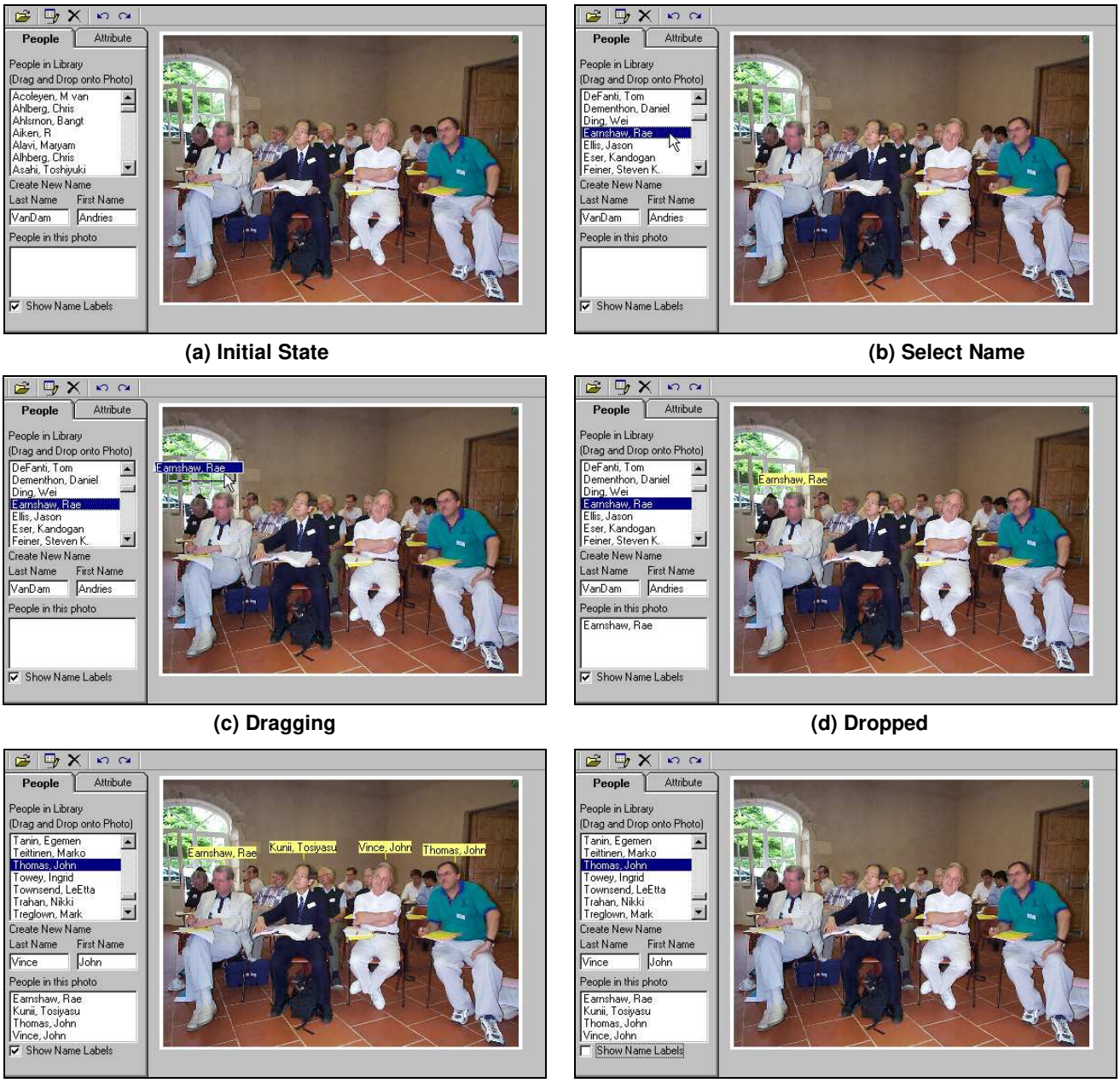

Figure 5. The Process of Dragging and Dropping an Annotation on a Photo

The selection list is shown as being an alphabetically organized scrolling menu, but it could be implemented as a split menu [13]. This would entail having 3-5 of the most commonly occurring names in a box, followed by the alphabetical presentation of the full list. Thus the most frequent names would be always visible to allow rapid selection. Name completion strategies for rapid table navigation would be useful in this application. When users mouse down on a name, the dragging begins and a colored box surrounds the name. When users mouse up, the name label is fixed in place, a tone is sounded, and the database entry of the XY coordinates is stored. The tone gives further feedback and reinforces the sense of accomplishment. Further reinforcement for annotation is given by subtly changing the border of photos in the Collection viewer. When a photo gets an annotation, its thumbnail's white border changes to green. Users will then be able to see how much they have accomplished and which photos are still in need of annotation.

A Show/Hide checkbox gives users control over seeing the photo with and without the name labels. Since the photo viewer window is resizable, the position of the labels changes to make sure they remain over the same

person. A small marker (ten pixels long) hangs down from the center of the label to allow precise placement when there are many people close together. The marker can be used to point at the head or body and it becomes especially useful in crowded group photos.

Future additions might include the capacity to resize the labels, change fonts, change colors, or add animations. Another interesting issue is collaborative annotation in which multiple users working side-by-side [14] or independently might annotate photos and then the results could be combined, with appropriate resolution of conflicts. Tools for finding variant spellings or switches between last and first names would help raise data quality. A valuable accelerator is bulk annotation [2], in which a group of photos is selected and then the same label is applied to every photo with one action, although individual placement might still be needed.

Of course, annotation by names of people in photos is only the first step. Drag and drop annotation for any kind of object in a photo (car, house, bicycle), map (cities, states, lakes), or painting (brushstroke, signature, feature) is possible. Annotation about the overall image, such as type of photo (portrait, group, landscape), map (highway, topographic, urban), or painting (impressionist, abstract, portrait) is possible. Colored ribbons or multiple star icons could be used to indicate the importance or quality of photos.

Searching and browsing become more effective once annotations are included in the photo database. The obvious task is to see all photos that include an individual. This has been implemented by simply dragging the name from the list into the Collection viewer or to a designated label area. The PhotoFinder finds and displays all photos in which that name appears in a label.

5. Database Design and Implementation

5.1 Schema of the Photo Library database

The PhotoFinder operates using a Photo Library database (Microsoft Access), which contains five linked tables (Figure 6). The basic concept is that a Photo Library contains Collections of Photos, and that Photos contain images of People.

In the Photo Library schema, the Collections table represents the collections of photos with attributes such as Collection Title, Description, Keywords, Starting Date, Ending Date, Location, Representative PhotoID and unique Collection ID. The Photos table is where references (full path and file name) of photos and their thumbnails are stored with important attributes such as the date of photo, event, keywords, location, rating, color, locale, and so on. Each photo should have a unique reference and photos with the same references are not allowed to be stored in this table even though they have different attribute values. The Linkage table is the connection between the Collections table and Photos table. It stores the links between collections and photos.

The People table stores all the information about the people who appear in the Photo Library. In our initial implementation, attributes include only the Given (First) name and Family (Last) name of the person, and the unique PersonID (people with the same first and last name are not allowed to be stored in People table). Eventually, the People table will be extended to include personal information such as e-mail address for exporting the Photo Library, homepage address, occupation and so on. The Appearance table stores the information about which Person is in which Photo. It serves as the linkage between the Photos table and the People table. Attributes include AppearanceID, PersonID, PhotoID, and relative (X, Y) coordinates (upper left corner is (0,0), lower right is (100,100)) of people in the photos.

In designing the Photo Library, we made three major assumptions concerning the Library, Collections, and Photos. These assumptions can be classified as follows:

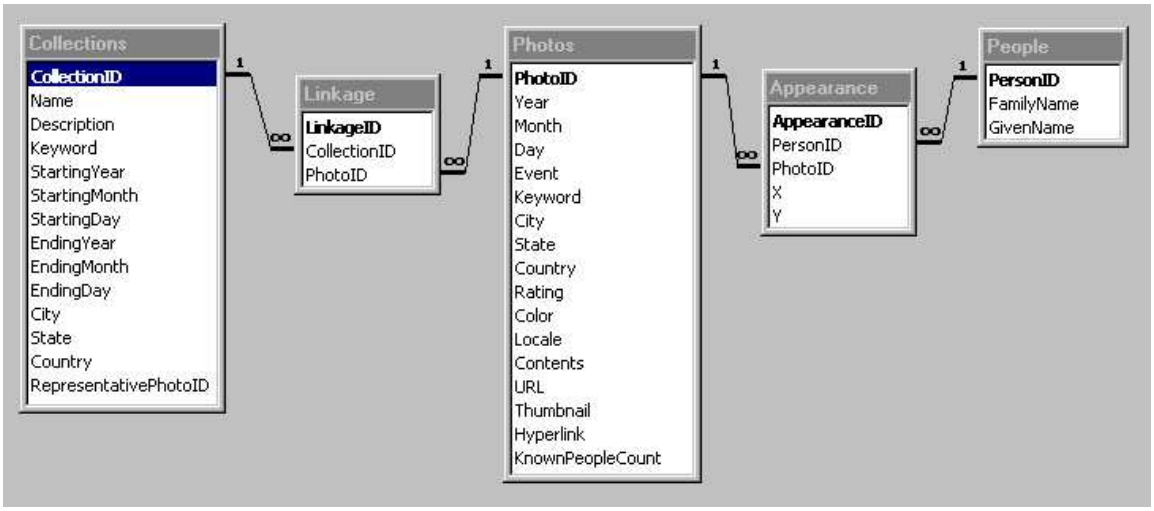

Figure 6. The schema of Photo Library database

- *Relationship between Collections and Photos*

A 1-to-many relationship between the Collections table and the Linkage table has been set so that a collection can contain multiple photos, and a 1-to-many relationship between the Photos table and the Linkage table has been set so that same photo can be included in multiple collections. It is also possible that a collection contains the same photo multiple times to permit reappearances in a slide presentation. Two different collections could have exactly same set of photos. If two photos have different path names, they are different photos even though they are copies of a photo.

- *Relationship between Photos and People*

A 1-to-many relationship between the Photos table and the Appearance table has been set so that a photo can contain multiple persons, and a 1-to-many relationship between People table and Appearance table has been set so that same person can be included in multiple photos. Multiple appearances of the same person in a photo are not allowed. A composite pair of Given name and Family name should be unique in the People table.

- *Relationship among Library, Collections, and Photos*

Within a library, the same photo could be contained in multiple collections multiple times, but their attributes and annotations must be the same.

In the first design of the Photo Library database, we only considered annotation by names of people in photos, but the Photo Library database can be easily extended by adding an Object table, Animal table, Keyword table, and so on, along with connection tables similar to the Appearance table. With such a Photo Library database design, more flexible annotation would be possible.

5.2 Updating the Photo Library Database by Direct Annotation

PhotoFinder keeps updating the Photo Library database whenever the direct annotation module causes any information changes. In this section, we classify the Photo Library database updating situations into five categories, and discuss corresponding algorithm and implementation issues.

- *Adding a New Name Label / Creating a New Person:*

When users drag a name from "People in Library" listbox and drop it onto a photo, PhotoFinder immediately checks whether there already exists an Appearance connection between the photo and the person since multiple appearances of the same person in a photo are not allowed. If a conflict occurs, PhotoFinder would highlight the existing name label on the photo and ignore the drag-and-drop event with a warning message. If there is no conflict, PhotoFinder finds the PersonID and PhotoID, calculates a relative (X, Y) position ($0 \leq X, Y \leq 100$) of the drag-and-drop point on the photo, and then creates a new Appearance record with this information. After adding a new record to the Appearance table, the PhotoFinder updates "People in this Photo" listbox and finally creates a name label on the photo. To show that the label has just been inserted, the newly added name in the "People in this Photo" listbox will be selected, and accordingly the new name label on the photo will be highlighted. If the added name label is the first one on the photo, PhotoFinder sends an event to the Collection Viewer to change the border color of the corresponding thumbnail to green, in order to show that the photo now

has an annotation. The algorithm for creating a new person is simple. As soon as users type in the first name and last name of a person in the editbox and press enter, PhotoFinder checks whether the name already exists in the People table. If so, a warning message will be displayed with the name in "People in Library" listbox being selected. If not, PhotoFinder creates and adds a new Person record to the People table, and then updates the "People in Library" listbox, selecting and highlighting the newly added name.

- *Deleting Name Label / Deleting Person:*

When the delete button of the Photo Viewer toolbar is clicked or the delete key is pressed, PhotoFinder checks whether the selected name label already exists. If not, PhotoFinder ignores the deleting action. But if it exists, PhotoFinder automatically calculates the PersonID of the selected name label and the PhotoID, and it searches through the Appearance table to find and delete an Appearance record having those IDs. PhotoFinder updates "People in this Photo" listbox and deletes the name label on the photo. If the deleted name label was the last one on the photo, PhotoFinder sends an event to the Collection Viewer to change the border color of the corresponding thumbnail to white, to show that the photo has no annotation. If focus is on the "People in Library" listbox and the delete key is pressed, PhotoFinder finds the PersonID of the selected name in the listbox. PhotoFinder deletes the PersonID from the People table and also deletes all the Appearance records containing that PersonID, which results in the complete elimination of the name label from the other photos in the Photo Library. Again, Collection Viewer updates the border color of thumbnails that no longer have annotations.

- *Editing a Name of Person:*

Users can edit a name of person in library by pressing the edit button of the Photo Viewer toolbar or by just double clicking over the selected name in the "People in Library" listbox. When the edited name is typed in, PhotoFinder finds and changes the corresponding person record from the People table only if there is no duplication of the name in the People table. It also refreshes both the "People in this Photo" and the "People in Library" listboxes, and all the name labels on the current photo. If duplication occurs, the whole editing process will be ignored with a warning message.

- *Positioning Name Label:*

Drag-and-dropping the existing label over the photo can change position of the name label. As mentioned before, the relative (X, Y) position of the center point of a name label is stored in the corresponding Appearance record. PhotoFinder uses a small marker hanging down from the center of the label to allow precise placement. But since the size and direction (downward) of the marker is fixed, it is somewhat difficult to distinguish labels when many people appear in the photo close together. Using Excentric labels [15] or adding an additional (X, Y) field to the Appearance table to allow a longer and directional marker could solve this problem. Other features such as changing the font size of labels and avoiding occlusion among labels in resizing the photo will be handled in future versions of PhotoFinder.

- *Importing People Table from other Libraries:*

Retyping the names that already exist in other libraries is very tedious and time consuming job. Therefore, PhotoFinder supports a function to import the People table from other libraries. The internal process of importing the People table is similar to that of creating a new person repeatedly. The only thing PhotoFinder should handle is checking and eliminating the duplication of a person name.

6. Conclusion

Digital photography is growing rapidly, and with it the need to organize, manage, annotate, browse and search growing libraries of photos. While numerous tools offer collection or album management, we believe that the addition of easy to use and enjoyable annotation techniques is an important contribution. After a single demonstration, most users understand direct annotation and are eager to use it. We are adding features, integrating search functions, and conducting usability tests.

Acknowledgements: We appreciate the partial support of Intel and Microsoft, and the contributions of Ben Bederson, Todd Carlough, Manav Kher, Catherine Plaisant, and other members of the Human-Computer Interaction Laboratory at the University of Maryland.

7. REFERENCES

[1] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey" *Proceedings of the IEEE*, Vol. 83, pp. 705-740, May 1995.

[2] Allan Kuchinsky, Celine Pering, Michael L. Creech, Dennis Freeze, Bill Serra, Jacek Gwizdka, "FotoFile: A Consumer Multimedia Organization and Retrieval System", *Proceedings of ACM CHI99 Conference on Human Factors in Computing Systems*, 496-503, 1999.

[3] E. Imhof, "Positioning Names on Maps", *The American Cartographer,* 2, 128-144, 1975.

[4] J. Christensen, J. Marks, and S. Shieber, "An Empirical Study Of Algorithms For Point-Feature Label Placement", *ACM Transactions on Graphics 14*, 3, 203-232, 1995.

[5] J. S. Doerschler and H. Freeman, "A Rule-Based System For Dense-Map Name Placement", *Communications Of The ACM* 35, 1, 68-79, 1992.

[6] John Lamping, Ramana Rao, and Peter Pirolli, "A Focus + Context Technique Based On Hyperbolic Geometry For Visualizing Large Hierarchies", *Proceedings of ACM CHI95 Conference on Human Factors in Computing Systems*, New York, 401-408, 1995.

[7] Jia Li., Catherine Plaisant, Ben Shneiderman, "Data Object and Label Placement for Information Abundant Visualizations" *Workshop on New Paradigms in Information Visualization and Manipulation (NPIV'98)*, ACM, New York, 41-48, 1998.

[8] Mark D. Pritt, "Method and Apparatus for The Placement of Annotations on A Display without Overlap", US Patent 5689717, 1997.

[9] Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price, "Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations", *Proceedings of ACM CHI 98 Conference on Human Factors in Computing Systems,* v.1 249-256, 1998.

[10] J Kelly Lee and Dana Whitney Wolcott, "Method of Customer Photoprint Annotation", US Patent 5757466, 1998.

[11] Richard Chalfen, *Snapshot Versions of Life*, Bowling Green State University Popular Press, Ohio, 1987.

[12] Ben Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction, 3rd Edition*, Addison Wesley Longman, Reading, MA, 1998.

[13] Andrew Sears and Ben Shneiderman, "Split Menus: Effectively Using Selection Frequency To Organize Menus", *ACM Transactions on Computer-Human Interaction 1*, 1, 27-51, 1994.

[14] J. Stewart, B. B. Bederson, & A. Druin, "Single Display Groupware: A Model for Co-Present Collaboration.", *Proceedings of ACM CHI99 Conference on Human Factors in Computing Systems*, 286-293, 1999.

[15] Jean-Daniel Fekete Catherine Plaisant, "Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization.", *Proceedings of ACM CHI99 Conference on Human Factors in Computing Systems*, 512-519, 1999.