

# 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](mailto: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 Photo Finder 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://www.qbic.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).

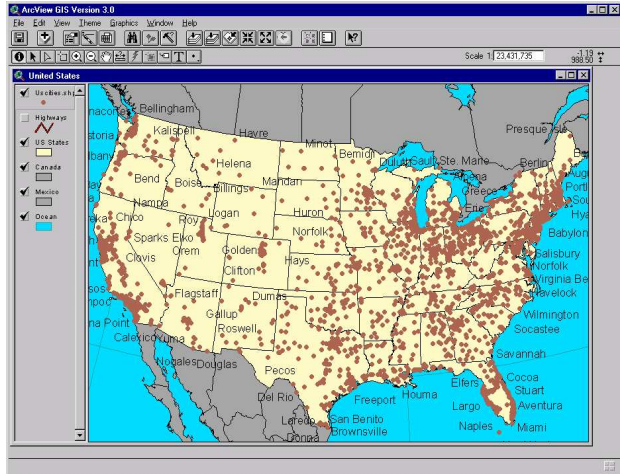


Figure 1. US Map with City Names

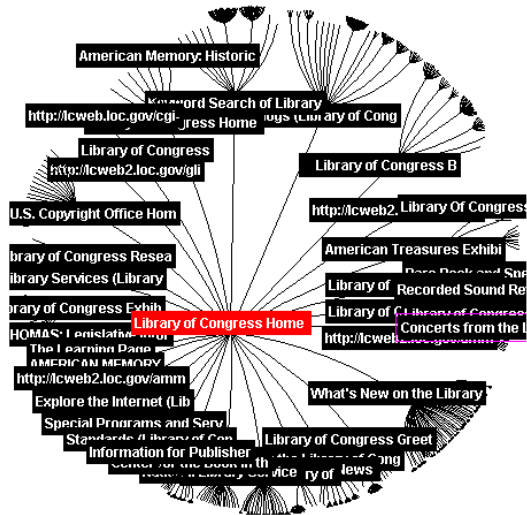


Figure 2. Hyperbolic Tree

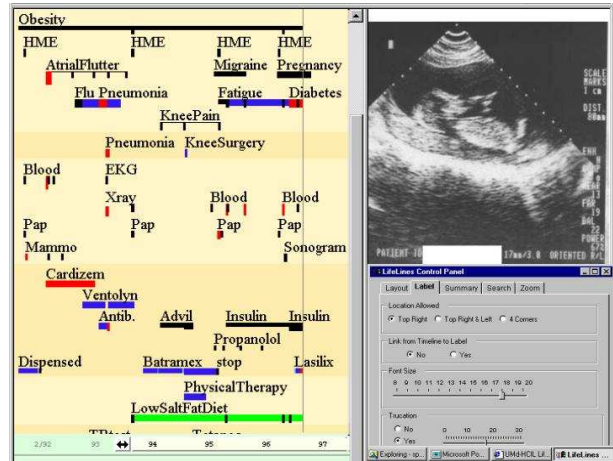


Figure 3. LifeLines Medical Patient History

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 interest 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 Photo Finder 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

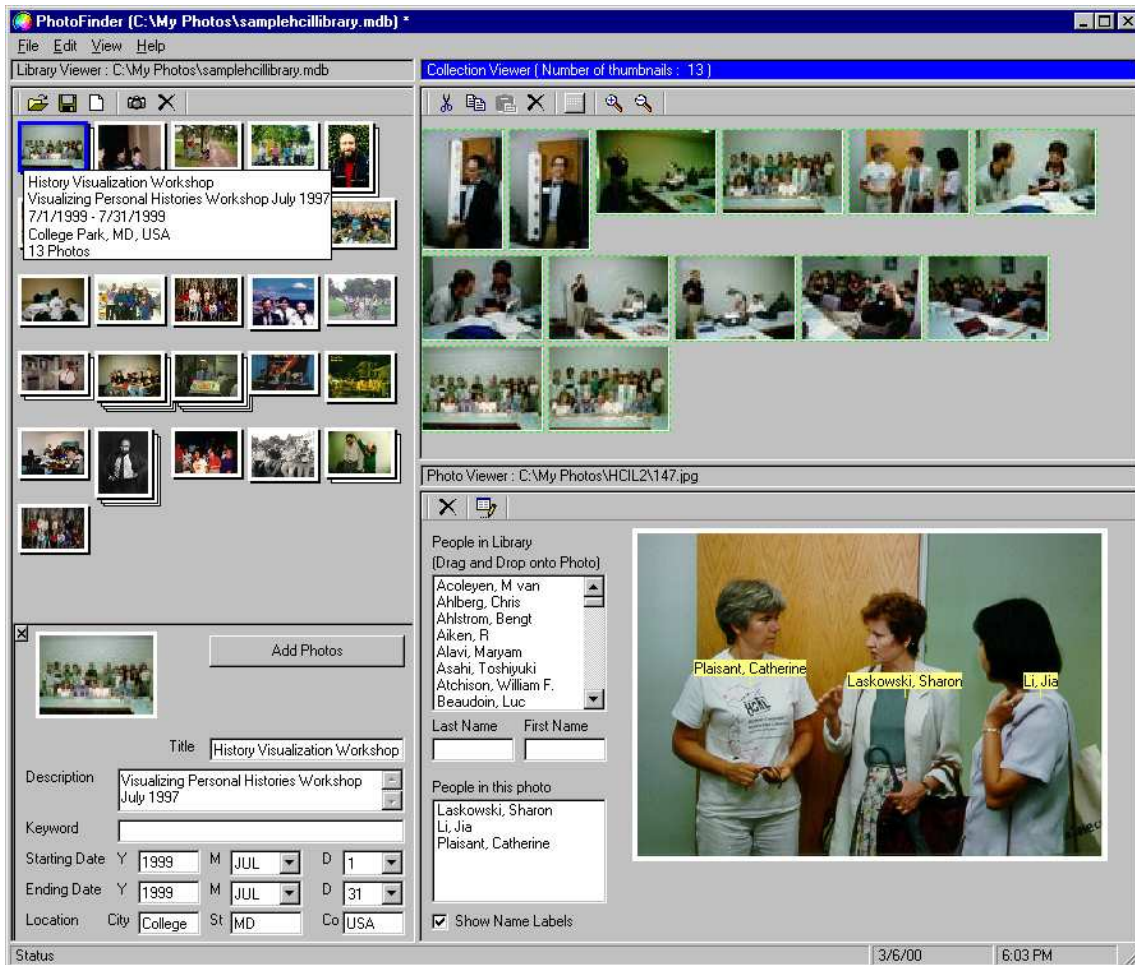
thumbnail designs were prototyped and will be refined for inclusion in future versions.

- **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 slideshow.

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.



**Figure4. PhotoFinder1 displaywith LibraryViewer ontheleft, CollectionViewerwiththumbnails ontheupperright,and PhotoVieweronthe lower right.**

#### 4. Direct Annotation

To cope with these challenges we developed the concept of direct annotation: selectable, draggable labels that can be placed directly on the photo. Users can select from a scrolling or pop-up list and drag by mouse or touchscreen. 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 simpler 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.



(a)InitialState



(b)SelectName



(c)Dragging



(d)Dropped



(e)FourIdentifiedPeople



(f)HideAnnotations

**Figure5.TheProcessofDraggingandDroppinganAnnotationonaPhoto**

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 mousedown 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

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

## LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

## FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

## E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.