Flexible Access to Photo Libraries via Time, Place, Tags, and Visual Features

Andreas Girgensohn, Frank Shipman, Thea Turner, Lynn Wilcox

FX Palo Alto Laboratory, Inc.
3400 Hillview Avenue
Palo Alto, CA 94304, USA

{andreasg, turner, wilcox}@fxpal.com

Department of Computer Science &
Center for the Study of Digital Libraries
Texas A&M University
College Station, TX 77843-3112, USA

shipman@cs.tamu.edu

ABSTRACT

Photo libraries are growing in quantity and size, requiring better support for locating desired photographs. MediaGLOW is an interactive visual workspace designed to address this concern. It uses attributes such as visual appearance, GPS locations, user-assigned tags, and dates to filter and group photos. An automatic layout algorithm positions photos with similar attributes near each other to support users in serendipitously finding multiple relevant photos. In addition, the system can explicitly select photos similar to specified photos. We conducted a user evaluation to determine the benefit provided by similarity layout and the relative advantages offered by the different layout similarity criteria and attribute filters. Study participants had to locate photos matching probe statements. In some tasks, participants were restricted to a single layout similarity criterion and filter option. Participants used multiple attributes to filter photos. Layout by similarity without additional filters turned out to be one of the most used strategies and was especially beneficial for geographical similarity. Lastly, the relative appropriateness of the single similarity criterion to the probe significantly affected retrieval performance.

Author Keywords

Photo libraries, photo retrieval, similarity criteria, visual similarity, geographic data, tagged photos.

Categories and Subject Descriptors

H5.1. Information interfaces and presentation: Multimedia information systems; H3.3. Information storage and retrieval: Information search and retrieval.

1. INTRODUCTION

Today, it is fairly common for users to have access to large photo collections. Finding photos of interest in such collections requires the use of different attributes of the photos. For example, when looking for photos of sunsets from Italy, one could filter by geographic location and then use a layout by visual similarity to place photos with shades of orange near each other. One could also use the time of the day to select among the Italy photos, assuming that the camera was set to the correct time zone. It would be even easier if those photos were tagged with “sunset.” We explore the benefit of similarity layout, the relative value of different attributes, and the value of being able to choose among them.

To let users make use of similarity layouts, we created an interactive visual workspace called MediaGLOW that presents a photo collection based on different similarity criteria. We currently offer four different similarity criteria: temporal, geographic, tag, visual. Temporal similarity is computed from the difference between photo creation times. Geographic similarity is based on the distance between latitude-longitude pairs. Tag similarity is computed using the Jaccard similarity coefficient of tags shared across photos. Our visual similarity is determined by an image classifier trained on manually tagged photos that compares predicted likelihoods for tags. In addition to grouping photos by similarity, MediaGLOW also provides three filters that restrict the time range, the geographic location, and the tags assigned to matching photos.

MediaGLOW integrates a variety of visualization and interaction techniques with different similarity criteria, enabling users to find relevant photos by proximity and by attribute filters. For placing photos in the 2D workspace, we chose a graph layout mechanism that visually indicates similarity among photos in the space while optimizing desired distances between photos. While grid-based layouts are more common for photo applications, they cannot accurately present similarity by proximity. Furthermore, while some similarity criteria, such as time, may naturally be visualized in one dimension, multi-dimensional similarity criteria can be visualized better in a two-dimensional layout.

We conducted a user evaluation to examine the hypothesis that a workspace grouped by multiple similarity criteria is beneficial for selecting photos. We also wanted to determine the utility of quickly filtering photos by different attributes and using photos to find more related photos. To that end, study participants were presented with photo collections retrieved from the Flickr photo sharing service [9]. The photos were selected based on their use of geographic locations and tags and were manually subsampled to make the collections equal size. Participants had to collect photos matching a set of probe statements. For some tasks, participants only had access to a single layout similarity criterion and filter option. For other tasks, MediaGLOW allowed the use of all similarity criteria and filters but started with different similarity criteria. Those variations provide insight into which photo attributes and system features are most useful.

In the next sections, we present a short scenario of use and discuss related work. We present MediaGLOW’s components, refined in several design iterations informed by user studies and informal observations. Next, we describe the setup of the user evaluation and the evaluation results. We conclude by discussing the photo data and system features that turned out to be most useful.

2. SCENARIO

As people have access to large collections of photographs, they need mechanisms for locating photos that match their current needs. The following scenario of finding photos for a travel brochure has similarities to many other photo selection tasks.

Joe is creating a travel brochure that will be used to promote tourism in California. As part of designing the brochure, photographs are needed that convey the diverse landscapes and activities found in the region. To appeal to as many people as possible, the photos should show different parts of the state at different times of the year. Joe has a large collection of photos of California, and uses MediaGLOW to select the best ones. Some of the photos he wants to include are the wine region in the fall, the mountains during winter, the desert during spring, and the beach during summer. To find pictures of the beach theme, he sets filters based on geographic location (Malibu) and time of year (summer) and views the results in the workspace. By filtering the collection by date and location, Joe reduces the number of photos being considered.

After selecting a photo of a surfer catching a wave, Joe makes use of MediaGLOW's layout that shows similar photos in close proximity. He switches the layout to visual similarity so that he can quickly see other photos having a similar blue color. He places a number of these photos into a stack for later selection. To find photos for the desert theme, Joe filters based on the tags "desert" and "cactus." He notices a photo of a cactus in bloom and creates a second stack with it. He then retrieves photos with tags similar to those in the new stack and adds the relevant ones. After creating a stack for each of the themes, he selects a photo from each stack such that the selected photos look good together.

This scenario is representative of collect-then-select processes that are found in a variety of contexts when working with photos. For example, when choosing photos for a web site, a slideshow, a wedding album, or to send to friends, candidate photos are selected from the broader collection. They are grouped based on the needs of the task (e.g., photos with particular relatives, friends, or activities) and then one or more of the candidate photos is chosen.

3. RELATED WORK

Work related to MediaGLOW falls into two main categories: research on access via alternative similarity measures and research on layout and interface techniques for interacting with collections. Similarity measures include visual similarity, access via tags and metadata, and the combination of several similarity measures. Layouts are frequently in the form of lists or grids. Non-rectangular presentations of data often use graph layout algorithms.

Access to photos via visual similarity is attractive as this can be computed in the absence of any metadata. Past research has not provided definitive answers regarding its utility for retrieving and grouping photos. Mills et al. [22] had somewhat negative conclusions regarding the utility of visual similarity. Later work in the same research group by Rodden et al. [27] found that grouping by similarity appears to be useful for designers. In our user evaluation, we attempted to determine the utility of visual similarity relative to other similarity criteria.

Photo access via tags and additional metadata is common. Flickr utilizes user-supplied tags and geographic tags to sort photos [9]. Photoshop Elements presents tag clouds or tags grouped by category [1]. MediaBrowser helps users annotate their photos by using time clustering and filtering [8]. Photo-Finder provides a drag-and-drop interface to simplify manual annotation [30]. PhotoSpread is a spreadsheet system for organizing photos using date, location, and other metadata [17]. The Rich Media Organizer [11] allows users to view their photo collections organized by time, location, or person. Faceted search and browsing interfaces generalize this concept to provide access through attributes in any order [32].

Ordering photos by time is one form of attribute-based access. For geo-tagged photos, it makes sense to place them on a map [16, 31]. While this metaphor is intuitive, it does not support the task-driven rearrangement of photos. Also, there are issues of scale mostly due to large gaps between photo locations. Given a location and time, additional metadata can be generated [23].

Related to our hypothesis concerning multiple similarity criteria, de Rooij et al. [28] show the benefit of offering multiple distance measures in searching video. From each displayed video shot, the user can branch to threads that include similar shots with respect to time, text similarity, and semantic similarity. They show superior performance over a system that just offers a temporal thread.

With regards to the layout of photos, most systems present thumbnails in lists or grids. Some applications allow sorting and access through photo attributes or automatically cluster photos based on such attributes. Still other applications provide graph-based or other non-rectangular presentations of photos.

Most commercial systems for browsing photos make use of a grid layout (e.g. iPhoto, Picassa, Photoshop Elements [15, 25, 1]). PhotoMesa provides a zoomable interface for browsing a grid-style layout of multiple directories of photos [2]. PhotoHelix [13] and TimeQuilt [14] provide interfaces for photo browsing that convey temporal order while making better use of screen space.

Automatic clustering of photos can aid user search. Systems such as Scatter/Gather [5] support narrowing a search by iteratively clustering the results and allowing the user to select relevant cluster for repeated steps. PhotoTOC automatically clusters photos using time and color [26]. A common task is the selection of a group of photos to share or publish as a group. Pixaura provides a manual interface for photo selection that supports tentative decision-making by allowing fuzzy membership in stacks [29].

Graph layout algorithms are frequently used for 2D layouts. Visual Who [6] uses a spring simulation to let group members' names gravitate towards anchors representing groups of people. The VIBE system [24] uses a force model to place retrieved documents near points of interest defined by key terms and a location. Dontcheva et al. [7] use a spring layout algorithm to indicate overlap in tags among photos returned via queries. Rodden et all [27] start with multi-dimensional scaling and use a greedy layout algorithm to place photos in a grid near their ideal positions.

Other studies [18, 21] focused on personal collections and found that people search only infrequently. However, search is useful for many photo collections, such as those in libraries, professional collections, photo-sharing sites, and other multi-user collections. To obtain quantitative performance results with photos with multiple attributes in our user evaluation, we selected several photo collections from Flickr [9] to be used by all study participants.

MediaGLOW combines attribute-based access, automatic clustering, and graph-based layouts. In contrast to previous work, MediaGLOW supports the iterative refinement and presentation of photo categories through a direct-manipulation, graph-based visualization. It includes some of the interactive machine learning tech-

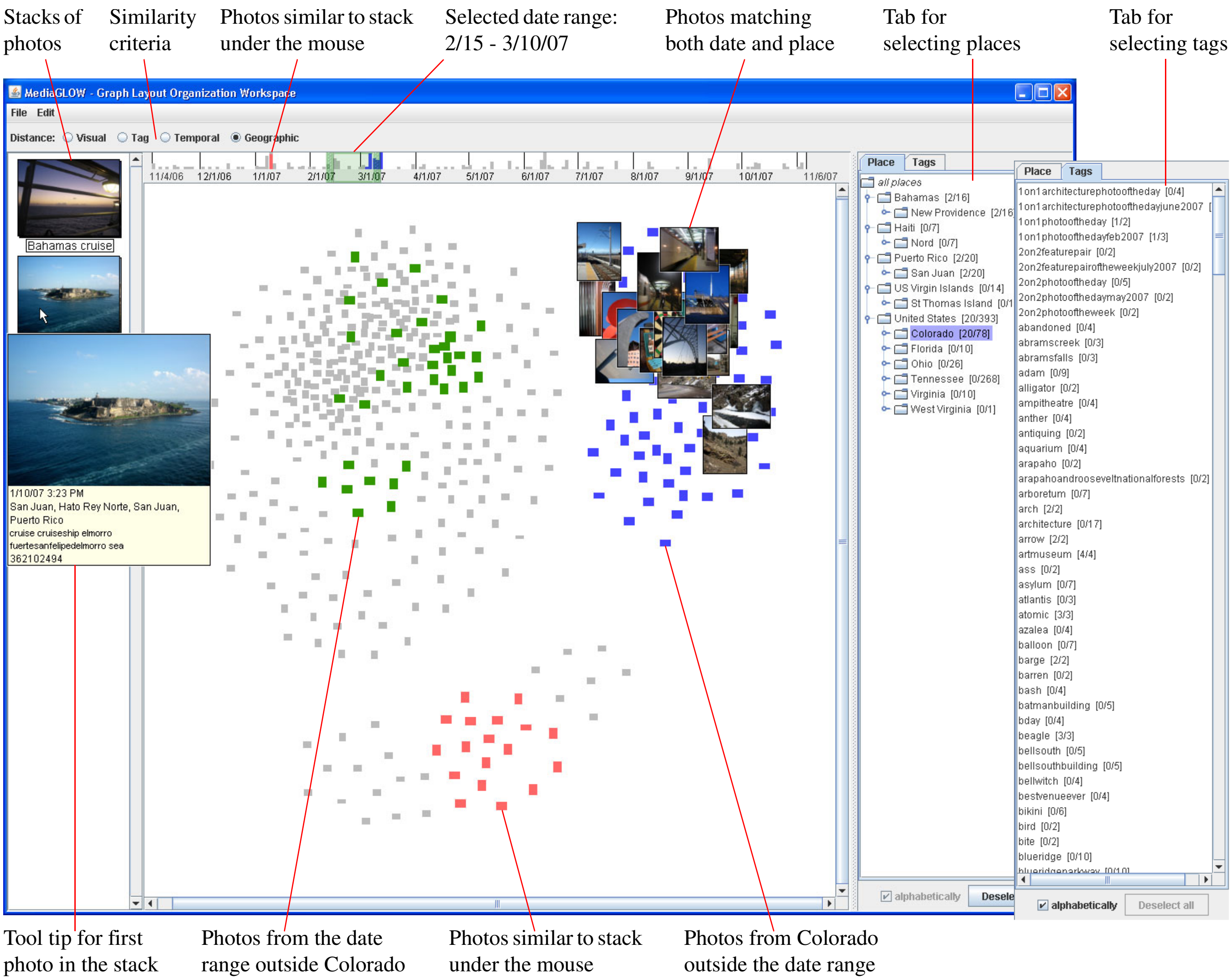


Figure 1. Photos Grouped by Geographic Similarity and Filtered by Date and Place.

niques from CueFlik [10]. It enables users to combine several grouping and filtering strategies to quickly find photos matching complex requirements. This makes MediaGLOW appropriate for exploring questions regarding the value of the different attributes and metadata in a photo collection. Furthermore, it supports situations where one or more of the attributes are appropriate.

4. PHOTO ORGANIZER WORKSPACE

We designed MediaGLOW with the hypothesis that grouping photos by several similarity criteria is beneficial for selecting photos. Instead of using a grid- or tree-based layout common in photo applications, we chose a two-dimensional, graph-based workspace to better present similarities amongst photos by proximity (see Figure 1). The workspace positions related photos near each other without being constrained to a grid. We offer multiple similarity criteria to support users in browsing related photos. Users can place photos into stacks to organize them. We also let users filter photos quickly by different attributes and use photos to find more related photos.

MediaGLOW has evolved over time. We initially developed a system for grouping photos solely by visual similarity. Over the past two years, we have experimented with many different attributes to help users better organize and navigate their photo libraries. We developed metaphors and controls to facilitate activities related to search, organization, filtering, and navigation of photos. Finally, results from a previous user study [12] contributed to the current version of MediaGLOW.

In the following subsections, we compare the various similarity criteria that MediaGLOW offers. We present interaction techniques incorporated into MediaGLOW to facilitate navigation in the workspace. Then, we describe the role of stacks and show how filtering and retrieval can help in selecting photos. Finally, we provide an overview of the workspace layout.

4.1 Similarity Criteria

MediaGLOW currently offers four similarity criteria for grouping photos: temporal, geographic, tag, visual. A simple but useful similarity criterion uses the photo creation times. If photos have location data, the geographic distance between the locations can be used to compute a similarity criterion. The haversine formula[1] is used to determine the distance between latitude-longitude pairs. The similarity between user-assigned tags is computed from the Jaccard similarity coefficient[2] of tags shared across photos. The vi-

[1] http://en.wikipedia.org/wiki/Haversine_formula

Figure 2. Expanded Selection With Photo Viewer.

sual similarity measure takes advantage of manually tagged data available at the Flickr photo sharing site. A classifier is trained on tagged photos to use visual attributes to predict the likelihoods for a set of 34 tags such as "mountain" or "summer." Those tags were selected as being common, meaningful, and expected to be visually discriminable. The vector of tag probabilities for two photos is used to compute the Kullback-Leibler distance [20].

Different similarity criteria produce different layout patterns in the workspace. For example, when organizing the workspace by a temporal similarity criterion, photos are grouped in a chain of tight clusters. The temporal distribution is due, in part, to people taking several photos in a row. Also, the temporal similarity is based on a single value so that placing photos along a straight line would be optimal. That line gets curved due to distance normalization, described later. Geographic similarity leads to clusters as shown in Figure 1. The visual and tag similarity criteria tend to evenly distribute photos (see Figure 5).

4.2 Navigating the Workspace

Photos can be selected in the workspace with standard selection gestures. Thumbnails of selected photos are slightly enlarged and the remaining thumbnails are faded out to emphasize a selection. On mousing over a photo, a tool tip appears showing a larger image and relevant information about that photo (see the left sides of Figures 1 and 3). Photos that have been filtered out by the user are shown as small, rectangular, grey or colored dots (see the center of Figure 1). Tool tips are available for those dots as well.

Exploring Sets of Photos

Selected photos in the workspace can be explored more closely in different ways. First, a circular tray displays thumbnails of all selection members arranged in one or more rings (see Figure 2). To make thumbnails fit better into the circular arrangement and for aesthetic reasons, the corners of the thumbnails are removed, leaving only ellipses. When the user moves the mouse over one of these ellipses, it is enlarged and the rectangular image is shown (see Figure 2). Second, a set of photos may also be viewed with a conventional viewer popup that allows the user to flip through those photos (see the left side of Figure 2).

Photo Stacks

Users may group photos into stacks of related photos as shown in the sidebar to the left of the workspace in Figure 1. Stacks are valuable to users because they enable the creation of and access to user-definable categories. In addition, stacks can be used by the system as examples for identifying photos similar to those already in the stacks.

Stacks act as implicit queries for photos similar to the photos they contain. Those photos are highlighted in the workspace and the timeline when mousing over a stack. A description of the retrieval of similar photos is below in the Section *Finding Similar Photos* (also see Figure 3).

Zooming in the Workspace

Users can zoom in or out of the workspace. If photo positions and thumbnail sizes were adjusted by the same zoom factor, users would not be able to separate overlapping photos. On the other hand, if the zoom factor were only applied to photo positions and the image sizes remained constant, users could not get a better look at the thumbnails. To address this issue, MediaGLOW includes two different zoom factors for photo positions and thumbnails. The cube root of the workspace zoom factor is used as the zoom factor for photo thumbnails. We found that it made navigating the workspace easier than our previous use of the square root because photos were spread out more quickly.

4.3 Filtering and Retrieving Photos

It is difficult to find particular photos amongst hundreds displayed in the workspace. Users may hide photos that do not match particular criteria. MediaGLOW enables filtering based on three of the four similarity criteria. Time can be restricted with sliders in the timeline (see the top of Figure 1). Geographic location and user-assigned tags are specified in tabs (see the right of Figure 1). Time can be combined with one of the other two. Filtering based on visual properties was left out due to the lack of well-defined filter boundaries. MediaGLOW can also select photos similar to a stack using any of the four similarity criteria.

Filtering Photos Matching Time or Attributes

The timeline (see Figure 1) shows a histogram indicating the number of photos taken on a particular day. A time range slider restricts the photo display to those taken within that time range. Color-coding of the histogram indicates time ranges with relevant photos.

A list of all tags assigned to photos can be used to show only photos that have selected tags. The tag list can be sorted alphabetically or by frequency to avoid having to deal with tags that match only a few photos. Geographic locations of photos are shown in a tree representing the hierarchy of locations (country, state, county, city). As with tags, selecting a location makes the corresponding photos visible. Both leaf nodes and intermediate nodes can be used for selections. Multiple selections are possible, e.g., selecting two countries to select all photos taken in either country.

Selections from the timeline and either the tag list or the location tree are combined such that only photos matching both are shown. The color-coded histogram in the timeline indicates times with

[2] http://en.wikipedia.org/wiki/Jaccard_index

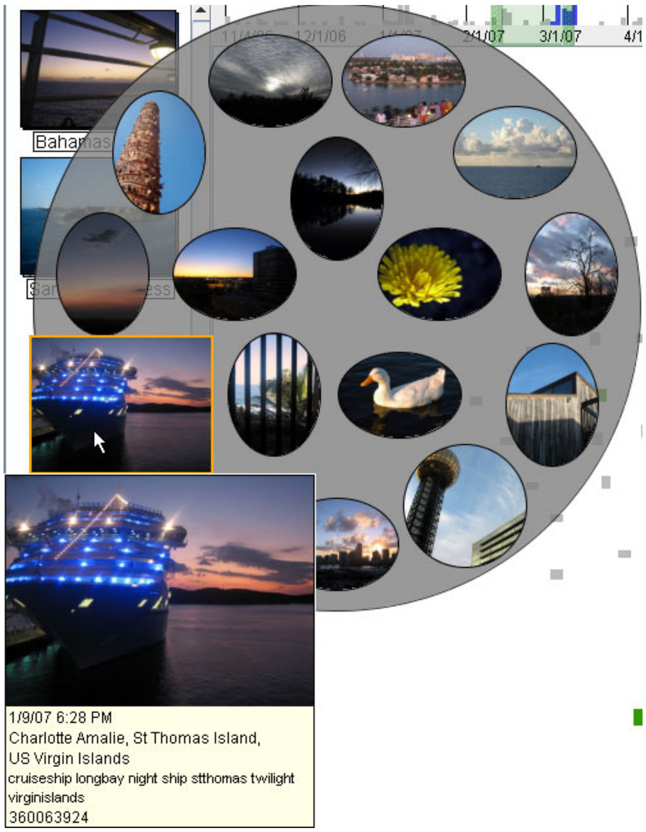


Figure 3. Retrieving Photos Visually Similar to a Stack.

photos matching tags or locations to support users in selecting the appropriate time ranges. In the workspace, small rectangular dots represent photos that do not match both criteria. Blue dots indicate tag or geographic matches, green dots indicate timeline matches, and gray dots match neither. Those colors were picked to match the colors of the timeline slider and of the selection in the attribute tags.

Finding Similar Photos

When hovering with the mouse over a stack, similar photos with respect to the selected similarity criterion are highlighted in the workspace and in the timeline (see pink dots and timeline bars in Figure 1). This provides the user with guidelines for where to explore to find those similar photos. Alternatively, the user can request a circular tray with photos similar to a stack where the most similar photos are in the center (see Figure 3). From this tray, the user can drag photos either into the stack used for finding those photos or into other stacks including newly created ones.

4.4 Graph-based Workspace Layout

MediaGLOW's layout supports user tasks by placing related photos near each other. Others have grouped photos while keeping them non-overlapping using tree maps [2] or self-organizing maps [19]. Rodden et al. [27] use a layout algorithm to place photos in a grid after the initial multidimensional scaling. However, non-overlapping presentations need more space. The 450 photos in Figure 5 would have to be scaled down from 68x51 to 28x21 to fit into a non-overlapping grid. While avoiding overlap is an important goal, the resulting organization does not necessarily keep related photos near each other. Furthermore, some of those approaches have computational requirements that prevent interactivity.

We chose to use a graph representing the similarity between photos. A spring model determines a layout in which the spring system is in a state of minimal energy. Such a graph layout can take placements by the user into consideration. This differentiates it from approaches such as multidimensional scaling [3]. Also, because only similarities are needed, a space with non-Euclidian distances can be used. In addition, by using a similarity metric with partial derivatives that guide the direction in which the iteration moves, a satisficing state can be found quickly.

In the following subsections, we compare approaches for layout frequencies, describe the force model for spring layout, and explain why we normalize spring lengths.

Dynamic Layout versus Workspace Consistency

In earlier versions of MediaGLOW, the layout of the spring network was recomputed each time the user moved a photo. While the dynamic layout provided a more accurate presentation of photo similarity, participants in a previous user study [12] had difficulties keeping track of which photos they had looked at. Creating independent subgraphs for clusters of similar photos provided greater stability but introduced additional issues. Independent subgraphs are fine as long as cluster membership is reasonably stable. However, users could modify stacks that formed the basis for those clusters such that subgraphs changed, leading to abrupt layout changes.

Instead of the above approach, the current system creates a layout once for every similarity criterion. Each similarity criterion is treated as a separate layer. Users can rearrange photos in each of those layers. When coming back to a similarity criterion, photos are where the user left them. The circular tray of similar photos described above replaces the functionality of photos drifting towards stacks of similar photos.

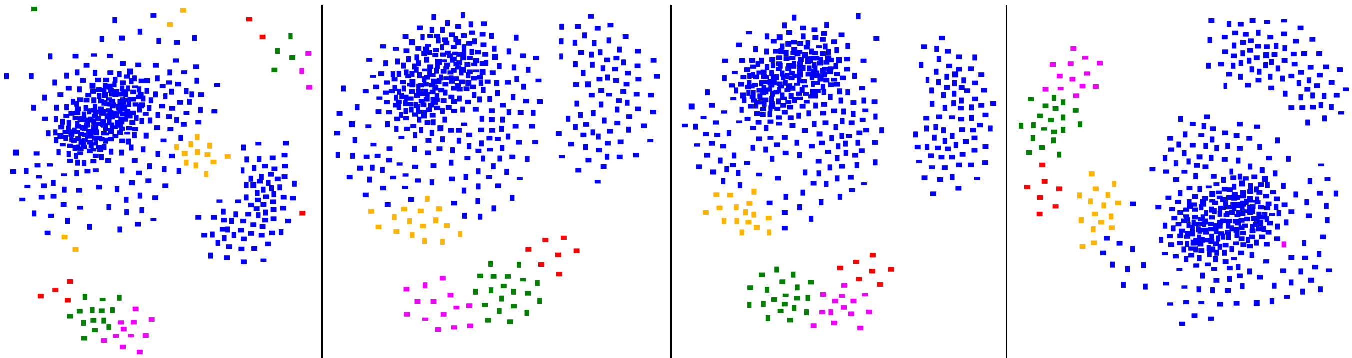
Spring Model

There are different approaches for creating a layout for the completely connected graph that represents similarities. MediaGLOW connects all pairs of graph nodes with springs to determine the layout of the photos. The main disadvantage of a spring layout compared to a grid layout is the overlap of photos. We address this issue in two ways. First, we assign a minimum length to all springs. Second, after the spring layout is completed, we push apart photos that are too close to each other.

The neutral lengths of the springs are determined by the particular similarity criterion that was chosen. Specifically, a spring is defined by a neutral length l and a constant k. Two nodes at distance d connected by a spring (l, k) are subject to the force $k * (d - l)$. This is a repulsive force if $d < l$. The length of the springs at rest corresponds to the desired distance between the nodes.

Using the same spring constant for all springs leads to unsatisfactory placements of some photos far away from similar photos as shown at the left of Figure 4. The repulsive force of longer springs overwhelms the attractive force of shorter springs. To address this issue, we made the spring constant k inversely related to the spring length l. We found that a quadratic relationship $k = 1 / l^2$ made it less likely for individual photos to be trapped in the middle of unrelated photos.

When placing photos in random positions at the start of the spring layout, many iterations may be required and the layout can get stuck in suboptimal states. Instead, MediaGLOW uses hierarchical agglomerative clustering to initially place photos. The cluster tree



Constant spring strengths; initialized with tree map; countries with low photo counts do not cluster well

Inverse distance as spring strength; good local clustering; global positions fairly good (Bahamas should be closest to Haiti)

Square inverse distance as spring strength; good local clustering; global positions not bad (Puerto Rico should be between Haiti and Virgin Islands)

Inverse distance as spring strength; random init.; best global positions (Bahamas closest to Haiti); one Virgin Islands photo trapped in the middle of the US

Blue: United States; yellow: Bahamas; red: Haiti; green; Puerto Rico; purple: Virgin Islands

Figure 4. Geographic Layouts with Different Spring Parameters.

is turned into a tree map where the sizes of subtrees determine relative areas in the tree map. Photos are placed inside the map areas. MediaGLOW then iteratively moves the photos in small increments towards a state of low energy.

Normalizing Spring Lengths

Similarity criteria may not lend themselves to be used directly as spring lengths. Both temporal and geographic distances have large gaps where no photos were taken. On the flip side, visual distances tend to cluster in the middle of the range. MediaGLOW normalizes distances by sorting all pairwise distances and by determining the percentile for each distance value. A distance at the 10 percentile of all distances would be assigned a spring length of 0.1 and a distance at the 80 percentile would be assigned a spring length of 0.8.

5. USER EVALUATION

We conducted a user study to determine the benefits of grouping by multiple similarity criteria and attribute filtering. Study participants had to locate photos matching probe statements such as "find pictures showing college campuses." We postulate that some similarity criteria are more suitable for some probes (i.e., retrieval tasks) than others. For example, sunsets can be found by visual similarity or by the time of day, photos in a given month can be retrieved by temporal similarity, and photos in a particular city can be found by geographic similarity. In the study, we evaluate the effects of placing similar photos near each other, letting users find more similar photos, and filtering photos by attribute. To evaluate the utility of these tools, we compare users' performance in locating photos matching a particular probe while either having access to all of MediaGLOW's tools or just a subset.

5.1 Study Materials

The Flickr service for managing and sharing photos provides programmatic access to a huge number of photos [9]. Many of the photos have been tagged by their owners and assigned to geographic locations, either manually or via GPS. We selected photo collections from Flickr that met certain criteria. We started with the complete collections of 3,000 owners that offered a large fraction of their collections under the Creative Commons license for use with attribution. This resulted in a pool of 11.8 million photos for which we downloaded metadata and images.

For selecting owners with geographically distributed collections, we divided the world into areas of 15 degrees latitude by 15 degrees longitude. We selected owners that took at least 50 photos in each of at least 10 different areas. Photos from the resulting 37 owners produced a pool of about 227,000 photos.

From the geographically distributed collections, we selected one-year subsets for 2006 and 2007 that had the most tags per photo. We subsampled those sets using only the photos with the highest number of tags while keeping the photos distributed across the year. Using this approach, we automatically selected sets of 500 photos that we manually pruned to 450. We selected 16 sets as suitable for the tasks in the study and three additional sets for training.[1] To distinguish between the use of tags and geographic locations in the study, we removed all geographic tags (e.g., Chicago) and temporal tags (e.g., April or 20060830) from the photos.

We created a list of probes such as "find photos of buildings in the summer sun" (see Table 1). We rated those probes with respect to the appropriateness of the four available similarity criteria and selected the probes with the most agreement among the authors. Appropriate and inappropriate similarity criteria were assigned to probes such that each similarity criterion is used four times as appropriate and four time as inappropriate. We then paired probes with photo sets such that there were at least 10 photos matching each probe, with an average of 35 matches. All participants were presented with the same 16 pairs of probes and photo sets, but in different orders and in different study conditions.

The resulting location tasks are similar to the collection activity from our scenario, where the user is selecting a set of candidate photographs about a particular theme from an online digital library of photos from others. This approach enabled us to use the same photo collections across study participants.

[1] We used photos from these Flickr users: (10297518, 12247055, 18432837, 21424326, 29549955, 30265340, 34589965, 38115734, 40279823, 45206671, 48096279, 61025927, 61577908, 63315403, 99357189)@N00.

#	Find pictures showing...	Appro-priate	Inappro-priate
1	religious art in Rome	*Geo*	*Visual*
2	glaciers	*Geo*	*Date*
3	college campuses	*Geo*	*Visual*
4	tall buildings in Chicago	*Geo*	*Tag*
5	music	*Tag*	*Visual*
6	rugby teams	*Tag*	*Date*
7	beer	*Tag*	*Date*
8	statues	*Tag*	*Geo*
9	christmas decorations	*Date*	*Geo*
10	buildings in the summer sun	*Date*	*Tag*
11	horse races in June	*Date*	*Tag*
12	railways in the autumn months	*Date*	*Visual*
13	snow	*Visual*	*Tag*
14	buildings near water	*Visual*	*Geo*
15	flowers	*Visual*	*Geo*
16	people in the water	*Visual*	*Date*

Table 1. Probes with Similarity Criteria.

5.2 Study Conditions

We tested four different *similarity criteria* for grouping and retrieving photos by similarity: *visual*, *tag*, *date*, and *geographic*. Participants either had an *interface* with just one of those measures (*single;* see Figure 5) or could switch among all (*full*). The similarity criterion presented when the application started was either *appropriate* or *inappropriate* for the probe. We hypothesized that the single-inappropriate condition would result in the worst performance. We also expected that starting with an inappropriate criterion would have less effect on performance with the full version of MediaGLOW, as participants could switch the similarity criterion.

In cases where participants were restricted to a single similarity criterion, we removed interface elements that would give them the effect of another similarity criterion. Those interface elements included the tabs for specifying geographic locations and tags shown at the right of Figure 1. If neither tab was available, the tab panel was empty (see Figure 5). The timeline was fully expanded and locked with the three non-temporal similarity criteria. If all similarity criteria were available, all other interface elements were available and the timeline was unlocked. The unlocked timeline started with about 40 days at the beginning of the date range.

5.3 Study Procedure

Participants. Thirteen male and three female participants took part in the study. Because of intellectual property concerns at the time of the study, all participants were employees of our research lab. All were familiar with photo browsing software.

Tasks. Each task consisted of a single probe and its photo set presented in either a single or the full interface, with a starting similarity criterion that was either appropriate or inappropriate. Participants were instructed to find as many photos matching the probe as possible in 2-1/2 minutes and place them in the top stack on the left. They were told that the time would not allow them to find all matches. The task duration was selected such that some photos could be found without permitting an exhaustive search or fatiguing the participants. The suitability of the task duration was verified in sessions with two pilot participants. After the completion of all tasks, each participant filled out an online survey. In that survey, 38% of the participants thought the time was too short and 44% thought it was about right.

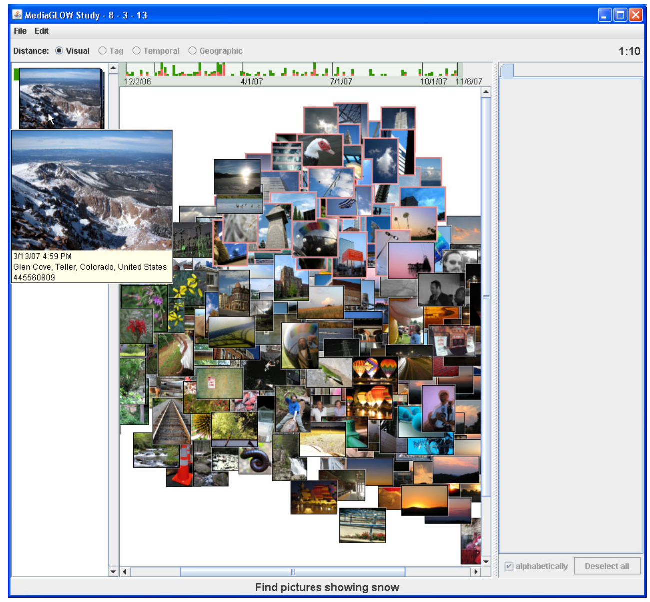


Figure 5. Study Interface With Only Visual Similarity.

Training. Each participant received about 40 minutes of training. First, they became familiar with the features of MediaGLOW and its use. Then they did five self-paced tasks with the same photo set used in the features training. The tasks used each of the four restricted, single similarity criterion versions, followed by the full version with all similarity criteria and features available. This was followed by two tasks with new photo sets and limited to 2-1/2 minutes each, one using the temporal similarity criterion only and the second the full version.

Balancing of Conditions. The study consisted of 16 2-1/2-minute tasks with 16 different probes, each with a different set of 450 photos, and with different combinations of study conditions and similarity criteria. We balanced similarity criteria and states both within each participant and across participants. Each participant saw each probe once in one of four possible conditions: single or full interface with a similarity criterion that was either appropriate or inappropriate to the probe. We also balanced combinations of similarity criterion and interface condition and of state and distance. No participant received a similarity criterion twice in a row. Finally, we balanced the average position of each probe across participants.

5.4 Study Results

To determine the effect of the similarity measures and system features, we examined retrieval performance, time spent in each similarity measure, and participants' stated preferences.

Recall, Precision and F-Score

To determine the ground truth for examining retrieval performance, we started with the union of all photos selected by participants as matches to a probe. Several of the authors rated these photos for relevance and discussed discrepancies. This approach for determining the ground truth is similar to that used at TRECVID. For measuring task performance, we use the F-score, the harmonic

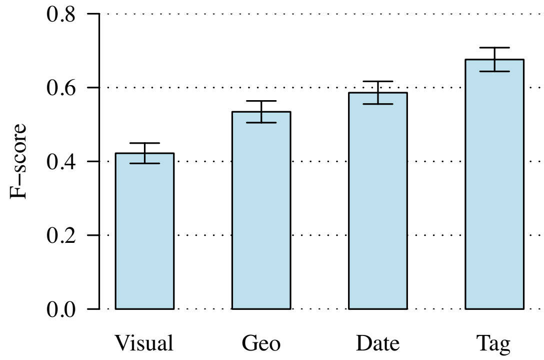


Figure 6. Retrieval performance by Similarity Criterion.

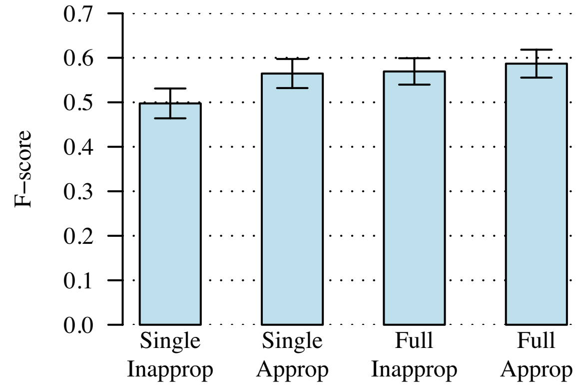


Figure 7. Users retrieved significantly fewer photos with inappropriate probes paired with a Single interface.

mean of precision and recall. Recall is the proportion of photos from the ground truth that were selected and precision is the proportion of the photos selected that were also in the ground truth. In the study, the average precision was 0.94 so that recall and F-score were almost the same.

An analysis of F-score showed that the similarity criterion used was an important determinant for finding matching photos ($F_{(3,45)} = 15.65$, $p < 0.001$). People were most successful using the Tag similarity criterion with its associated tab and least successful using the Visual criterion. Date and Geographic were between them (see Figure 6). In the Single interface, the Tag criterion performed better than the alternative six out of eight times. However, in two probes where we rated it as inappropriate, that condition performed much better than the one selected as appropriate (probes 11 and 13 in Table 1). The interaction of similarity criterion and appropriateness was significant ($F_{(3,45)} = 12.03$, $p < 0.001$), largely due to the strength of tags in the two inappropriate cases. In addition, performance was better in the Full interface than in the Single interface ($F_{(1,15)} = 4.45$, $p < 0.052$).

We predicted that the Single interface with an Inappropriate probe would show the poorest performance (see Figure 7). For tasks using a Single interface in an Inappropriate similarity criterion, participants were forced to use what was available. However, for tasks starting with an inappropriate configuration using the Full interface, participants could — and did — switch to more appropriate similarity criteria and filters. To test this, we did a planned orthogonal comparison between the Single/Inappropriate condition and the other three. Indeed, people had the most difficulty finding matching photos in the Single/Inappropriate condition ($F_{(3,36)} = 4.67$, $p < 0.05$). The other conditions did not differ significantly from each other. In each of these three conditions, participants either started with the most appropriate similarity criteria or were able to change to it. Only in the Single/Inappropriate condition did they have to find matches using an inappropriate criterion.

Methods for Retrieving Photos

People used a variety of methods to restrict displayed photos in their search for matches, with varying success. We categorized methods as follows:

- *Similar:* "Show Similar" used to obtain a subset,
- *Layout:* similarity criteria for arranging layout, no filters,
- *Date:* restricted timeline,
- *Tag:* selected tag(s),
- *Geo:* selected geographic location(s),
- *Date&Tag:* restricted timeline and selected tag(s),
- *Date&Geo:* restricted timeline and selected location(s).

Many retrieval methods could not be used in tasks presented with a single criterion interface. Therefore we examined the use of the methods separately for the single and full interfaces. Table 2 shows the total number of retrieved photos by retrieval method (columns), interface condition, and the similarity criterion selected at retrieval time (rows). Methods unavailable to single criteria interfaces have no retrievals.

For single interfaces, retrieval was not significantly different for similarity criteria or method due to wide variation in retrieval rates, but the interaction between them was significant ($F_{(12,280)} = 6.57$, $p < 0.0001$). In Visual, with no available filters, the preferred method was *Similar*, accounting for 61% of all matches in Single-Visual (154 of 251 in Table 2). Use of the filter methods *Date*, *Tag*, and *Geo* dominated their respective single criterion interfaces, while *Layout* alone was used to retrieve 27% of single interface matches (470 of 1751).

For full interfaces, the ability to use different criteria and filters allowed the full range of methods, including combinations of criteria. As a result, we observed differences in the type of methods that were used for retrieving matches ($F_{(6,392)} = 3.84$, $p < 0.001$). Of the five methods used in the single interface, all but *Tag* decreased in use, as people employed the combination methods of *Date&Tag* and, to a lesser extent, *Geo&Tag*. As expected, methods interacted with similarity criteria ($F_{(18,392)} = 2.24$, $p < 0.01$), as some methods could only be used with certain criteria. Participants gravitated towards *Date&Tag* and used *Tag* more in the Date and Geo Layout; these combinations were not available to them in the single interface. Showing the strength of multiple criteria, the timeline was increasingly used, but in combination with tags or geographic locations. The utility of the similarity-based layout was demonstrated by its strong retrieval results both in single and full interfaces that were only exceeded by tag selection.

Using Different Similarity Criteria

In half of the tasks, participants used the full version of MediaGLOW. This allowed them to change the similarity criteria at will. They could make full use of the features of MediaGLOW to change the time frame displayed, to select geographic or tags as filters, and to use combinations to combine visual, tag, date and geographic attributes. In the restricted versions, participants did not need to spend time considering which similarity criteria would be most effective or conceptualizing more complex strategies. For

Interface		Similar	Layout	Date	Tag	Geo	Date& Tag	Date& Geo	Total
Single	Visual	154	97	-	-	-	-	-	*251*
	Tag	21	127	-	471	-	-	-	*619*
	Date	7	139	241	-	-	-	-	*387*
	Geo	43	107	-	-	344	-	-	*494*
	Total	*225*	*470*	*241*	*471*	*344*	-	-	*1751*
Full	Visual	11	54	21	35	10	90	9	*230*
	Tag	1	97	37	318	2	285	54	*794*
	Date	3	53	25	61	15	199	15	*371*
	Geo	15	154	17	98	82	92	121	*579*
	Total	*30*	*358*	*100*	*512*	*109*	*666*	*199*	*1974*
Grand total		*255*	*828*	*341*	*983*	*453*	*666*	*199*	*3725*

Table 2. Total Number of Photos Retrieved.

tasks in the Single interface, participants spent a total of 300 seconds in each similarity criterion.

In the full version, because they could change between them, the amount of time spent in each is a measure of their preferences for the criteria. All but one participant made use of the ability to change criteria in tasks when using the full version. The Tag criterion was used most often (379.0 s), followed by Geo (313.5 s). Visual and Time were used the least (256.5 s and 251.1 s, respectively). The difference in time reflects how useful the arrangement of photos by criteria was judged to be ($F_{(18,359)} = 6.80$, $p < 0.001$).

Post-Task Questionnaire

The stated comfort with search modes and the perceived effective strategies matched the performance discussed above. The full version with all similarity criteria was preferred and the visual-only mode was seen as ineffective. Most participants were happy with the features provided by MediaGLOW. 75% answered "yes" or "sure" when asked whether they would like a system with MediaGLOW's features and the rest provided more guarded answers such as "maybe" or stated that they would like to see those features assist more traditional photo applications.

When asked what they liked about MediaGLOW, 75% of the participants indicated that they liked the ability to control the different criteria used for searching, viewing, and sorting photos. As one said, "Options, options, options. I really like being able to control how I see and search for pictures." Another participant indicated that he liked the "multi-faceted organization" and elaborated that "different tasks call for different approaches." Most participants wanted to use strategies combining several criteria, that it was "nice to be able to search using several dimensions." Another liked the "very diverse search criteria and browsing styles."

A third of the participants commented favorably on the search by similarity feature. It allowed them to search using criteria that are hard to express in words." The interface itself had mixed reviews, ranging from "fun to zoom in and out of pictures" to "hard to see all the images in the workspace." 25% of the participants would like to see a traditional grid layout as part of MediaGLOW.

5.5 Discussion

From the results above, we conclude that it is important to give people multiple ways to organize and search photo collections. This may seem obvious, but many systems are designed with a presentation organized around a single form of access, be it through tags, time or geography. While prior studies show success with time and event-based access, our study shows that people's strategies are heavily influenced by the available features, as seen in Table 2. While strategies such as filtering by tags were frequently used, participants selected photos from the workspace without using filters 22% of the time (828 of 3725 in Table 2), making this the second most common selection technique. While the workspace layout was frequently used, it also led to comments indicating that some participants would prefer non-overlapping views at least for part of the task. This is something that needs to be addressed in future system versions. More generally, users made use of combined methods by using attribute filters while the workspace layout was based on a different similarity criterion. As expected, restricting participants to tools inappropriate for the task had significant detrimental effects on their performance.

In comparing the different access methods, tags were the most successful and preferred criteria of the four we tested. That criteria's success, however, depends on the quality of the tags provided in the collection. Flickr photos tend to have many and meaningful tags. Furthermore, we selected collections that had large numbers of tags in addition to distributed geographic locations. For our study collections, most photos had three or more tags. Although some tags were inaccurate or not useful, people frequently used both the tag similarity criteria for organizing the collection in the workspace and tags in combination with time and geographic information. While the photo collections in the study were real, our selection process resulted in a high tag-to-photo ratio and tags are not always available in personal collections. Useful tags are difficult to generate automatically. Hence, while tags are the most preferred source of information in the metadata studied, they are the least likely to be available without significant up-front effort.

Date and Geographic location worked well and are likely to be generated by cameras. The geographic workspace layout had the most distinct grouping, so that participants did not use filters as much with it. Date filters were used by themselves or in combination with other filters. The temporal layout of the workspace was used the least, probably because the timeline slider offers very similar functionality.

Less useful was the Visual similarity criterion, the only option available for photo collections lacking any metadata. A quarter of our probes were well aligned with the visual properties of the photos. However, even in these cases, people often switched to other criteria if given the choice. When filters were not available, visual similarity search was frequently used and led to reasonable results. This is a clear indication that the performance and understandability of visual search has to be improved to become truly usable.

6. CONCLUSIONS

We presented MediaGLOW, an interactive workspace that combines visual metaphors and user interface techniques with similarity-based 2D-layout and filtering techniques. MediaGLOW provides access to photo collections through a variety of similarity criteria. It goes beyond traditional browsing interfaces for photo organization by allowing users to find relevant photos using a variety of approaches.

We conducted a user evaluation to determine the utility of similarity layout and to compare the utility of different similarity criteria and filtering methods. Study participants frequently made use of the similarity layout to select photos from the unfiltered workspace grouped by different criteria. In addition, users also made heavy use of filtering by various attributes and found photos using all criteria, alone and in combination, whenever they were available. Not surprisingly, many photos were found using tags (alone and in combination with other criteria) assigned by the photo owners as they carry the most semantic meaning. Nonetheless, strategies using other criteria were equally successful. When restricted to a similarity criterion considered inappropriate for a task, participants performed significantly worse.

Finally, the results support our approach of offering multiple means to find photos. Users made use of the various similarity-based layouts and multiple filters to select desired photos. The study demonstrated that multiple similarity criteria in a combination of layout and filtering can support a wide range of selection tasks and collections. Because photograph libraries cannot predict all of the tasks that their patrons are likely to have, interfaces like MediaGLOW support access to and increase the value of these libraries for the general population.

7. REFERENCES

1. Adobe Photoshop Elements. http://www.adobe.com
2. Bederson, B.B. (2001). PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. *Proc. ACM UIST '01*, pp. 71-80.
3. Borg, I., Groenen, P. (2005). Modern Multidimensional Scaling: theory and applications (2nd ed.), New York: Springer.
4. Chang, M., Leggett, J., Furuta, R., Kerne, A., Williams, J.P., Burns, S., Bias, R. (2004). Collection Understanding, *Proc. JCDL '04*, pp. 334-342.
5. Cutting, D., Karger, D., Pedersen, J., Tukey, J. (1992). Scatter/Gather: a cluster-based approach to browsing large document collections. *Proc. ACM SIGIR '92*, pp. 318-329.
6. Donath, J.S. (1995). Visual Who: Animating the Affinitiese and Activities of an Electronic Community. *Proc. ACM Multimedia '95*, pp. 99-107.
7. Dontcheva M., Agrawala M., Cohen M. (2005). Metadata Visualization for Image Browsing. *UIST '05 Demonstration*.
8. Drucker, S. M., Wong, C., Roseway, A., Glenner, S., De Mar, S. (2004). MediaBrowser: reclaiming the shoebox. *Proc. Advanced Visual Interfaces '04*, pp. 433-436.
9. Flickr. http://www.flickr.com/
10. Fogarty, J., Tan, D., Kapoor, A., Winder, S. (2008). CueFlik: Interactive Concept Learning in Image Search. *Proc. ACM CHI '08,* pp. 29-38.
11. Girgensohn, A., Adcock, J., Cooper, M., Foote, J., Wilcox, L. (2003). Simplifying the Management of Large Photo Collections. *Proc. INTERACT '03*, pp. 196-203.
12. Girgensohn, A., Shipman, F., Wilcox, L., Turner, T., Cooper, M. (2009). MediaGLOW: Organizing Photos in a Graph-based Workspace. *Proc. IUI '09,* pp. 419-424.
13. Hilliges, O., Baur. D, Butz, A. (2007). Photohelix: Browsing, Sorting and Sharing Digital Photo Collections. *TABLETOP '07,* pp. 87-94.
14. Huynh, D., Drucker, S., Baudisch, P., Wong, C. (2005). Time Quilt: Scaling up Zoomable Photo Browsers for Large, Unstructured Photo Collections. *CHI '05 Extended Abstracts*, pp. 1937-1940.
15. iPhoto. http://www.apple.com/photo
16. Jaffe, A., Naaman, M., Tassa, T., Davis, M. (2006). Generating Summaries for Large Collections of Geo-Referenced Photographs. *Proc. Multimedia Information Retrieval*, 853-854.
17. Kandel, S., Paepcke, A., Theobald, M., Garcia-Molina, H., Abelson, E. (2008). Photospread: a spreadsheet for managing photos. *Proc. ACM CHI '08*, pp. 1749-1758.
18. Kirk, D., Sellen, A., Rother, C., Wood, K. (2006). Understanding Photowork. *Proc. ACM CHI '06,* pp. 761-770.
19. Kohonen, T. (2001). Self-Organizing Maps. Berlin: Springer.
20. Kullback, S. (1987). The Kullback-Leibler distance. *The American Statistician* 41, pp. 340-341.
21. Miller, A., Edwards, W.K. (2007). Give and Take: A Study of Consumer Photo-Sharing Culture and Practice. *Proc. ACM CHI '07,* pp. 347-356.
22. Mills, T., Pye, D., Sinclair, D., Wood, K. (2000). Shoebox: A digital photo management system. *Technical Report 2000.10.* AT&T Laboratories Cambridge, UK.
23. Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H., Paepcke, A. (2004). Context Data in GeoReferenced Digital Photo Collections. *Proc. ACM Multimedia 2004,* pp. 196-203.
24. Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., Williams, J.G. (1993). Visualization of a Document Collection: the VIBE System, *Information Processing & Management*, 29(1), pp. 69-82.
25. Picassa. http://www.picassa.net
26. Platt, J.C., Czerwinski, M., Field, B.A. (2003). PhotoTOC: automatic clustering for browsing personal photographs. *Information, Communications and Signal Processing*, 6-10.
27. Rodden, K., Basalaj, W., Sinclair, D., Wood, K. (2001). Does Organisation by Similarity Assist Image Browsing? *Proc. ACM CHI '01,* pp. 190-197.
28. de Rooij, O., Snoek, C.G.M., Worring, M. (2008). Balancing Thread Based Navigation for Targeted Video Search. *Conference On Image And Video Retrieval*, pp. 485-494.
29. Satpathy, L., Kamppari, S., Lewis, B., Rhee, Y.W., Elgart, B., Prasad, A., Drucker, S. (2008). Pixaura: Supporting Tentative Decision Making when Selecting and Sharing Digital Photos. *Proc. HCI '08*, pp. 87-91.
30. Shneiderman, B., Kang, H. (2000). Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. *Proc. Int. Conf. Information Visualisation (IV2000)*, pp. 88-95.
31. Toyama, K., Logan, R., Roseway, A. (2003). Geographic Location Tags on Digital Images. *Proc. ACM Multimedia 2003,* pp. 156-166.
32. Yee, P., Swearingen, K., Li, K., Hearst, M. (2003). Faceted Metadata for Image Search and Browsing. *Proc. ACM CHI '03*, pp. 401-408.