Adventures in Space and Time: Browsing Personal Collections of Geo-Referenced Digital Photographs

Mor Naaman, Susumu Harada, QianYing Wang†, Andreas Paepcke
Stanford University
mor, harada, paepcke@cs.stanford.edu, †wangqy@stanford.edu

ABSTRACT

We evaluate two novel applications for browsing personal collections of geo-referenced digital photographs. The first, PhotoCompas, is a browser that employs no graphical user interface elements other than the photos themselves (textual browser). PhotoCompas was developed in our project, and is based on an automated organization of the respective photo collection into clustered locations and events. The second application, WWMX, features a richly visual interface, which includes a map and a timeline. WWMX is a third party implementation. We conducted a user study, where subjects performed tasks over their own geo-referenced photo collections. We found that even though the participants enjoyed the visual richness of the map browser, they surprisingly performed as well with the textual browser as with the richer visual alternative. This result argues for a hybrid approach, but it also encourages textual user interface designs where maps are not a good choice. For example, maps are of limited feasibility on handheld devices, which are candidates for replacing the traditional photo wallet.

Categories and Subject Descriptors

H.5.1 [**Information Systems Applications**]: Information Interfaces and Presentation—*Multimedia Information Systems*; H.5.2 [**Information Systems Applications**]: Information Interfaces and Presentation—*User Interfaces*

General Terms

Human Factors, Algorithms

Keywords

Photo browser, geo-referenced photos, GPS, personal photo collection

1. INTRODUCTION

How would consumer photographers most profitably interact with collections of their own digital photographs, if each photo were stamped with both the time when the image was shot, and the geographic coordinates where the picture was taken? We explore two possible answers: a visually rich map-based interface, and a textual interface used to access a collection-specific, automatically created time/location hierarchy. The former system relies on the widespread familiarity with geographic tools, the latter appeals to a more analytic mental model.

Managing personal collections of digital photos is an increasingly difficult task. Not only do collections grow as years go by, but the rate of digital acquisition rises as storage becomes cheaper and "snapping" new pictures gets easier.

Current approaches to the photo collection problem include tools that encourage manual annotations, techniques for accelerating the visual scanning of photographs on the screen, and image analysis techniques. We discuss some of these systems in Section 6.

Our approach has been to deploy automatically collected metadata in place of human generated annotations. We utilize this metadata to support the browsing and searching of collections, hopefully into the tens of thousands of photos. For example, in [7, 8] we show how the *timestamp* embedded by digital cameras in every photo file can be used for organization that does not require any human effort, but is useful nonetheless.

Technology advancements have made it feasible for cameras to add location information to digital photographs as well. This information consists of the exact coordinates where each photo was taken[1]. Location is one of the strong memory cues humans rely on when they recall past events: the triplet of *who*, *where*, and *when* is generally thought to be the predominant memory aid [22].

Location information is therefore an important characteristic of photographs that invites exploration. We in fact aim to develop applications that build on the location *as well* as the time metadata that is already embedded in digital photos.

As a foundation for this work we set out to determine experimentally how well two very different applications enable end users to utilize the two cues in concert. The first system emphasizes an analytic model [14], the second relies primarily on visual techniques that display geographic maps [20].

Maps are an obvious way to communicate and manipulate location information. We hypothesize that maps are a

[1]There are a number of ways to produce "geo-referenced photos" using today's off-the-shelf products. For a summary, see [20].

powerful tool and will be useful for most users. However, maps may be impractical in some important situations, as they are inefficient in their use of screen real estate. For example, the map based overview of a photo collection that comprises photos from just San Francisco and Paris would occupy much of the screen with (visual) geographic information that is not pertinent to the collection.

The problem intensifies when the user operates on a small-screen device. Such devices seem well positioned to replace the traditional harmonica display of photographs that stuffs many travelers' wallets. Yet, maps are not likely to be well suited for this environment. Textual browsers are more likely capable of screen estate parsimony. Moreover, limited input mechanisms (such as cell phone inputs or voice activation, for example) may not be well suited to map-based manipulations. Finally, a number of people are uncomfortable with maps and prefer other types of browsing.

On the other hand, a non-map interface must provide intuitive structure and names for the different location as handles for the user to manipulate. Even with a good organization and recognizable geographic names, it is hard for such an interface to support the same level of detail as a map.

We conducted a controlled experiment to determine how well users understand and operate the map and non-map metaphors. The first system we studied is PhotoCompas ("PhC"), a system we developed. PhC [14] is based on automatically generated hierarchical time and location categorization of the photos, using thin textual menus navigation. The only graphic elements of the interface are the photos themselves. Figure 1 shows one of PhotoCompas' screens.

The second system we studied is the World Wide Media Exchange (WWMX) [20]. The WWMX system has a photo browsing interface that is based on an interactive map and timeline, using a map and other strong visual elements. A sample WWMX screen is shown in Figure 3.

Note that in addition to location, both systems require absolutely no human effort in organizing the photos. Also, note that both systems use time and the powerful cross-filtering of time and location for browsing the photo collection.

Of course, a combination of both systems is possible and likely beneficial as different metaphors could be used at different times. However, the current sharp contrast between the two systems allows for more focused evaluation of each approach's strength, and user preferences.

The rest of this writing is organized as follows. In the subsequent Section 2 we describe in more detail the two systems we studied. In Section 3 we describe the experiment's procedure. Results and analysis are offered in sections 4 and 5.

2. TWO EXPERIMENTAL BROWSERS

As we developed the PhotoCompas system, we describe it here in more detail. We briefly describe the WWMX application as well so the reader will be presented with the full picture regarding our two experimental systems.

2.1 PhotoCompas

Our system, PhotoCompas, is a hierarchy-based photo management system that despite the lack of a map supports efficient browsing of a personal library of geo-referenced photos.

PhotoCompas ("PhC") is comprised of two parts. The first part, which we developed, performs the computation

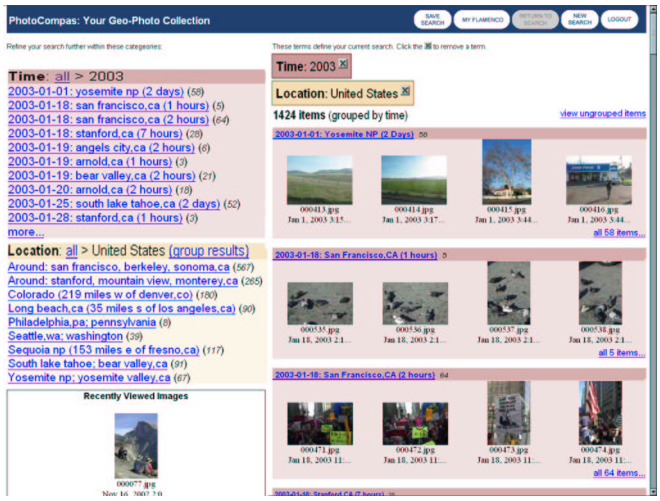

Figure 1: Screen shot of PhotoCompas. This view shows the collection restricted to photos taken in the United States during 2003. The events under 2003, and locations under United States are shown.

required to automatically group the user's photos into distinct events and locations, and assigns the location and event proper geographical names. Figure 2 shows a subset of a structure that was created by the PhC processing for a sample collection, including node names as they were assigned by PhC. The dashed lines represent parts of the location hierarchy that exist but have been abbreviated for the purpose of the figure, although there are more missing nodes than the dashed lines seen in the figure.

The second part of PhC is the interface. We use PhC's automatically generated structure in a "plug-able" interface, namely the Flamenco interface from Yee et al [23]. Flamenco is designed to make use of hierarchical faceted metadata, making it a good complement to our system. We describe the interface here first; in Section 2.1.2 we expand on the underlying algorithms.

2.1.1 *PhotoCompas Interface*

We used the Flamenco HTML-based toolkit [23] to create the interface for PhotoCompas. The authors of Flamenco kindly made their implementation of a general metadata browser available to us. That system allowed us to prototype a browsing interface quickly, so that we can proceed to build a complete user interface once we have the certainty that users understand and are able to navigate PhC's generated structure. Previously, Flamenco has been used to provide access to art collections, medical records, and similar applications.

In the Flamenco interface toolkit, users navigate the photo collection by clicking on different categories (i.e. nodes) in the hierarchy of each dimension. Such dimensions are called *facets*. In the Flamenco/PhotoCompas installation, as shown in Figure 1, there are two available hierarchical facets: the aforementioned Location and Time/Event hierarchies. The facets appear on the left hand side of the screen.

The Location facet portrays the location hierarchy as it was created by PhC's processing, using the automatically generated textual captions as described below. An simple

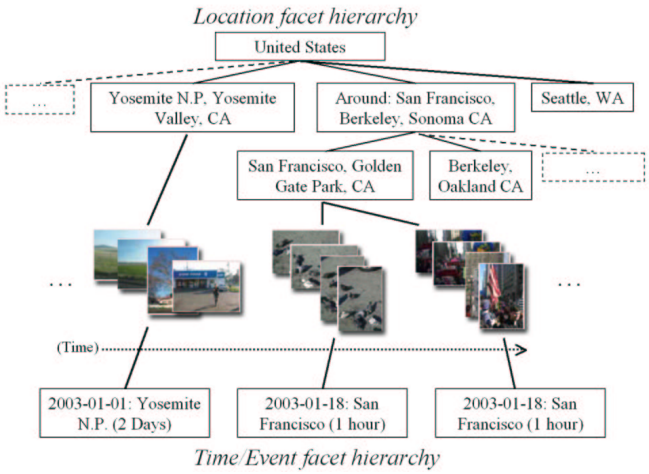

Figure 2: Sample PhotoCompas structure. Parts of the location and time/event hierarchies for an actual collection of photos.

example of a location hierarchy is shown at the top of Figure 2. The top nodes in the Location hierarchy are countries. The next level is the breakdown of the country into lower-level location categories. In Figure 1, for example, we show a subset of the location hierarchy that is presented after the user drilled into "United States" photos in our sample collection. Under the Location facet bar (on the left hand side of the figure), we see the nodes that are descendants of "United States", e.g. "Yosemite N.P.; Yosemite Valley, CA". In this interface, clicking on a node in the hierarchy also constrains the photo collection to only show photos belonging to this node. The photos are restricted to "United States" at this point; If we click into Yosemite, the collection will be further restricted, and the descendants of the Yosemite node (if any) will show under the Location facet.

For the Time/Event facet, the automatically created structure is flat — simply a sequence of events as detected by PhC's event analysis. The sequence is shown at the bottom of Figure 2. However, the full list of events can be long and would therefore be difficult to browse. Instead, we place the events into the time hierarchy by creating a "year" level. Each event can be found under the year in which the event started. For example, a path in the Time/Event hierarchy could be "2003"→"2003-03-12 San Francisco (2 Hours)." The event name includes the start date and duration, as shown in Figure 2.

Indeed, in Figure 1, the user had already clicked on the year "2003" in the Time/Event hierarchy. The set of photos that is displayed in the Figure's snapshot is therefore constrained both by the location ("United States", as seen above) and the time ("2003"). Consequently, under the Time categories we only see those events that happened in the United States during 2003. In addition, the US locations that appear in the figure are *only* the locations the user had visited during 2003.

Users can easily alternate between facets when they navigate the collection. A possible navigation path is "United States" (location facet), "2003" (Time/Event facet), and "Yosemite N.P." (location again). At that point, the user is presented with all the photos that are part of the Yosemite location, taken at events that occurred in 2003. The interface allows the user to quickly drill in and out of the collection using both facet hierarchies. In addition, the pictures that match the current filter (the current selection of categories) can be grouped by any other facet; e.g. users can click to have the "Sri Lanka" photos organized by event to quickly find the photos from a certain event during their trip. In Figure 1, for example, the displayed photos are grouped by time/event, each event represented by the first four images.

In addition to the Location and Time/Event facets, we have explored other types of metadata that can be automatically derived using the time and geographic coordinates of the photos [13]. In particular, given exact time and place where a picture was taken, we are able to use a number of data sources to deduce the actual local time; the daylight status (was it night, day, sunset or sunrise); and even the weather conditions and the temperature at the time the picture was taken. This additional metadata enables, for example, browsing for a photo that was taken in England, on a rainy day, just before the sun would have set.

We could have had these metadata available in the experiment's PhC interface, but decided to leave it out as it was not available in the WWMX interface, and would thus have given PhC a result-distorting advantage. Instead, we will evaluate the usefulness of the additional metadata in future work.

Next, we show how the grouping by location and event was created for the interface by PhC's processing.

2.1.2 PhotoCompas Collection Processing

To review, the main goal of PhotoCompas' automatic collection processing is to group the user's photos in two dimensions, The first dimension is location, and the second is event. That is, we wish to group the photos into hierarchies of locations and time-based events. Naturally, these two dimensions interact: photos from a certain event are associated with a location; any location may have pictures taken in it at different times (for example, multiple trips to Yosemite National Park).

The location/event derivation is complex. In the case of location, we have to represent the different areas the user visited in a way that feels natural and intuitive for navigation. Only if locations are intuitive, and furthermore, are automatically named effectively, can textual navigation be successful.

Events are equally difficult to identify, as the human notion of "event" varies: on one hand, multiple picture-taking days can be a considered a single event (e.g., a trip to New York); on the other hand, photos taken in a single day may be thought of as different events: a birthday event closely followed by an unrelated dinner party, for example.

While the full details about the algorithm appear in [14], we provide some detail here for completeness.

First, note that although both time and location can easily be subjected to some pre-defined hierarchy, PhC does *not* make use of this hierarchy when creating the grouping. For example, any specific time can be easily categorized into a year, month, day, hour hierarchy. Similarly, a location can be categorized by country, state, county, city etc. While these existing hierarchies are convenient, they are not optimal for the purpose of organizing photos.

The main argument against using a pre-defined hierar-

chy is that it is often too rigid, or placement of a given time/event in the hierarchy is ambiguous. A simple example that illustrates the problem is an event that straddles a day's boundary. Similarly, a road trip across multiple States did not "occur" in any one of the traveled States. We discuss this issue of appropriate hierarchies further in [14].

PhotoCompas processes a photo collection in two steps. The first step partitions the photographs into groups that will "make sense" to the collection's owner. Both time and space enter into this first step, using techniques of clustering and segmentation. The second step attempts to derive descriptive names for each group of photos (each node in the location and event hierarchies) by mapping the coordinates of each photo to a named location such as a city or a park, and generating an appropriate name for each. We first describe the hierarchy-generating steps, and then the naming algorithm.

The location and event hierarchies are created simultaneously. Initially, PhC sorts the photos by time. PhC then treats the photos as a sequence, and looks at the time and geographical distance between each pair of consecutive photos to create an initial grouping into *segments* representing low-level events. In other words, the photo sequence is initially segmented such that a breakpoint is created whenever a statistically irregular gap in space and time is detected between two consecutive photos.

Given the initial segmentation, PhC clusters these segments using a purely geographic clustering algorithm. The algorithm assumes that each segment occurs in a specific geographic location, and tries to find a minimal set of such locations that still enables a good mapping from each segment to a location.

The result of this step is a single-level location hierarchy, where each of the initially identified segments is assigned to a location in the hierarchy. At this point, some locations may be overloaded with segments. For example, the location where the photographer lives and takes most of her photos is likely to contain many photos across a wide time range. A remote vacation spot that the photographer visited once or twice would be sparsely populated with segments.

Based on criteria such as the number of photo-taking days in a single location, PhC may further split a location into lower-level clusters, and re-assign the respective segments to each new cluster.

Finally, PhC makes another time-based pass over the sequence of photos to decide the final breakdown into events, informed by the newly created location clusters. Briefly, the event breakdown is done by first creating event boundaries where the location grouping changes, and using adaptive threshold which is based on number of different visits to each location to create event boundaries for consecutive events in the same location.

Once this first step is completed, PhC needs a way to present these results in a user interface, without the benefit of a map. The second processing step is therefore to assign textual names to the nodes in the location and event hierarchies. The names for the nodes are based on the locations where photos belonging to these nodes occur.[2]

Our naming process consists of three steps. In the first step, we find for each latitude/longitude pair the state, city, and/or park that contain that location. This containment analysis uses an off-the-shelf geographic dataset of administrative regions.[3] For example, a particular coordinate may be inside in California (state), San Francisco (city), and Golden Gate National Recreational Area (park). Another coordinate may occur in Washington (state) and Seattle (city), but not in any park.

We count the frequency at which each named region occur in the set of photo coordinates, building a term-frequency table. We weigh each entity differently, as the different weights allow us to give more importance to names that are more likely to be recognizable to users (e.g., city names). At the end of this process, we have a *containment table* with terms and their score.

In the second naming step, we look for neighboring cities. By locating cities that are close to the coordinates in this set, and computing the distance from the center of the set to the city, we are able to produce textual names for these clusters such as "40 KMs south of San Francisco". We pick neighboring cities based on their "gravity": a combination of population size, the city's "Google count", and (inversely) the city's distance from the center of the set of photos. The "Google count" of a city is the number of results that are returned by Google [6] when the name of the city (together with the State) is used as a search term. We use this as a measure of how well known a city is and thus how useful it would be as a reference point. This step creates a *nearby-cities table*, again with terms and their scores.

The final step involves picking 1–3 terms from the containment table and the nearby-cities table to appear in the text caption of each set of photo coordinates (whether it is a location or event node). For example, a possible caption can include the two top terms from the containment table, and the top nearby city: "Stanford, Butano State Park, 40 KMs South of San Francisco, CA". Our method of picking the final terms is different for events and locations. The details can be found in [14]. Generally speaking, for locations, PhC uses the top term for the containment table, unless the term does not have a high enough score; in this case more terms will be added from the containment and nearby tables. Event names are based solely on the top term in the containment table, and are augmented by the starting date and duration of the event.

2.2 WWMX

The World Wide Media Exchange (WWMX) [20] is a state-of-the-art map-based application for digital photos. This application was originally designed by Toyama et al. for a *global* collection of geo-referenced photos (thus "*World Wide* Media Exchange"). Still, WWMX's photo browser user interface is also designed to be used for a single user photo collection. While the application implementation and interface are described in detail in [20], we summarize it here briefly for completeness.

The WWMX browser uses a powerful map and timeline interface. At any point during the browsing process, the user can view the map, the timeline and a set of photos that occur within both the boundaries of the displayed map and the limits of the displayed timeline. For example, if a map of the United States is shown, and the timeline is set

[2]Of course, events usually represent some context that is more elaborate than location, yet unavailable to our processing, e.g. "Mom's Birthday".

[3]Regretfully, we only have access to a database of US cities and parks. Thus, we have only tested our naming procedure on US photos.

Figure 3: Screen shot of WWMX. The view is restricted to photos taken in the United States during 2003.

to display the year 2003 only, the photos displayed will be ones that match both these filters. A corresponding sample screen shot of the WWMX application is shown in Figure 3.

The photos are shown as thumbnails in a dedicated photos pane, but also as dots on the map and the timeline. The dots are consolidated into larger dots when they occur in proximity. Users can pan and zoom the map to show a different area and therefore a different set of photos. An efficient way to do so is to draw a rectangle over the map; the map then zooms into the rectangle. Similarly, users can pan and zoom the timeline to restrict the display to photos from a specific time.

While the interaction in both the location and the time dimensions is different than the PhC interaction, the effect of the interaction is in essence equivalent: each interaction focuses on either the location dimension, or the time dimension. Therefore, we decided that comparing WWMX and PhC will enable us to evaluate the benefits and drawbacks of a text-based interface and a map implementation. The following section describes the experiment we devised for this purpose.

3. EXPERIMENT PROCEDURE

The goal of the experiment was to determine how subjects use both systems to perform (i) focused browsing/search for specific photos, and (ii) browsing for photos based on a loose theme. For each subject, we used the subject's *own personal collection* of photographs. To evaluate, we looked both for subjective and objective measures: from task completion time and mouse clicks to a questionnaire about preferences and attitudes.

The experiment followed a within-subject design. We exposed each subject to two experimental conditions: the PhotoCompas browser and the WWMX map-based browser.

Each subject completed two tasks on each browser. The first was a Search Task. We showed the subjects one of their own photos on a computer monitor and asked them to find that photograph in their collection by navigating the application. We set no time limit for this task, but timed it and asked subjects to work as efficiently as they could.

In order to minimize experimenter bias during the selection of photos for the Search Task, we had a computer randomly select the photos from each subject's collection. The computer presented one random photo after another to one of the experimenters. The experimenter accepted or rejected each photo based on the following criteria: a photo was rejected if (1) the picture was taken at the same event as one that had previously been chosen, or (2) the photo did not display any recognizable context, and the subject would not have been able to identify the photo in the collection. All other photos were accepted. The study in [15] followed a similar procedure and reports positive experience with this approach. We used this procedure in [8], also with positive results.

The second task was a Browsing Task. We asked the subject to select "good" pictures for a collage that represented some portion of the subject's life. We randomly alternated the collage topic between the two conditions. The two collage topic choices were "friends and family", and "trips". We asked subjects to select photos from as broad a time span and set of occasions as possible. We did not impose a time limit, but rather asked them to "stop when you feel you found enough photos" (which usually took 4–5 minutes).

For each browser we had subjects complete the photo Search Task six times, having each subject find a different photo each time. We asked subjects to perform the Browsing Task once for each browser. Each time subjects completed both tasks under one of the conditions, they were asked to complete a questionnaire. We asked questions such as the helpfulness of the photo organization, the subject's degree of satisfaction, the amount of frustration, and adequacy of the allowed time. Answers were encoded on a 10-point Likert scale. We also timed the Search Task, measured the number of mouse clicks for the Search and Browse task, and counted the number of pictures found during the Browse task.

3.1 Recruiting Participants

It is practically impossible to find any geo-referenced personal photo collections today, as the geo-photo technology is not available to many consumers yet. We solved this problem by using a "location stamping" tool. The Location Stamper[4] allows users to retrospectively mark their photos with a location, by dragging and dropping their photos onto a map. Thus, our subject pool was extended to all users with a collection of digital photos, even if those photos are not initially geo-referenced. This relaxation of participant selection constraints, then, required our participants to invest extra time in location stamping their photos. This activity required 1.5 hrs per user, on average.

Even with the location-stamping tool, since digital cameras have only become popular in the last few years, it was difficult to find participants with "regular" sizeable photo collections. A further constraint was that participants needed to feel comfortable about giving us access to their photos.

At the end of the search process, we were able to recruit 15 participants for our experiment; one was used as the "pilot" study, and another experienced severe technical problems and was ruled out. We report below the results for the remaining 13 participants.

[4]Available from http://wwmx.org

3.2 Statistics and Setup

Participant ages ranged from 17 to 49, with the highest representation in the 20s. Five of our participants were male, and eight were female. As both PhC and WWMX utilize time as well as location information, we did not process photos that had missing time or location information. The average processed collection size was 1,489 images. The average time span of the collections was 2.5 years. On average, each collection contained photos from 3.2 different countries, and had 80% of the photos taken in the United States.

For the experiment, we loaded the subject's photo collection onto a desktop PC with 512MB of RAM, dual 2.8GHz Intel Pentium 4 processors, and a 21" flat panel display with a resolution of 1280x1024 (WxH) pixels and 32bit colors. The thumbnails we generated for the photos in PhC were 140 pixels long on their longer edge. The WWMX application used a smaller thumbnail size that fit inside a rectangle of 42x30 (WxH) pixels.

3.3 Other Procedural Considerations

There are a few limiting yet unavoidable issues regarding the experiment design. The Location Stamper seems to advantage the WWMX interface, as participants are already exposed to the location of their photos on the map when they stamp their photos in the first step of the experiment. When the subjects later use the WWMX interface, they are slightly biased as they were exposed, and even created, the map locations of their photos. For this reason, we tried to have at least a few days between each user's location-stamping session to the experimental session. However, due to time and participant's personal constraints we could not ensure such a gap for all participants.

On the other hand, WWMX was disadvantaged because we used "manual" referencing of photos, rather than using an accurate location-capturing device. As the referencing accuracy was determined by each user when they were using the stamping tool, we did not have the full-scale accuracy of a GPS — especially when users were referencing trip-related photos, from locations they did not know as well. Arguably, the added accuracy may benefit a map-based application. It must be noted that the manual referencing may have hurt PhC as well — photos were often marked in locations that did not allow PhC to pick the best name for each set of photos. To give an example, one of the subjects had all photos of a certain set located at some point outside a well known park. Our naming algorithm, as described above, did not pick that park's name since none of the photos was marked as having been taken within the park itself, even though in reality some photos were shot within the park's borders. More accurate GPS based location acquisition would have improved PhC's naming performance.

Some deficiency of PhC may have influenced the results. The Flamenco interface toolkit, while being the correct choice for the fast prototyping facility we required, is limited to HTML-based interactions. Even a mostly textual interface can be visually optimized for a particular application. HTML does not provide the necessary control for such optimization.

In addition, we could not employ the techniques we developed [7] for selecting summary sets of photos to represent an event or location; the Flamenco interface only displays the first few photos of a set, and does not currently allow the application to participate in selecting which photos will be displayed to represent a given collection of items.

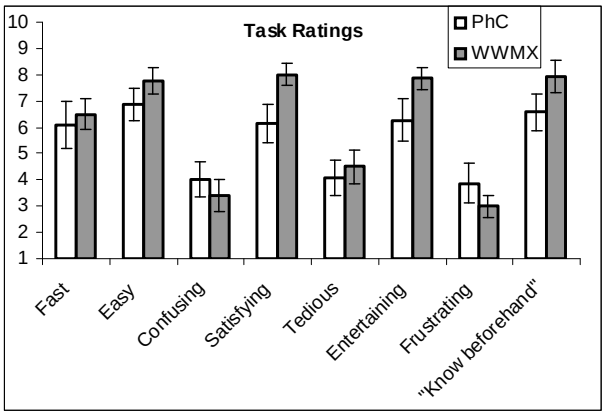

Figure 4: User subjective evaluation of the tasks in both applications.

Finally, it must be noted interaction with these two systems is not entirely independent. For example, if a participant was exposed to PhC first, they may during the first part of the experiment become familiar, or at least be reminded, with the content of their collection. As a consequence, interacting with the second system may be aided by this knowledge. To this end, we took care to randomize the order in which participants were exposed to the two conditions.

4. RESULTS

When we speak of quantities being 'the same' in this and the following Sections, we mean that the difference between the quantities was not statistically significant: $p > 0.05$. Measures we call 'different,' 'better than,' etc., correspondingly manifested differences at $p < 0.05$.

Figures 4–6 show some of the objective and subjective measured results. The task ratings of Figure 4 refers to subjects' responses to how they felt about performing the two tasks on either of the interfaces. The ratings were given on a 1–10 Leikert scale. A higher rating for a certain keyword means subjects found the keyword more appropriate for the particular interfaces. For example, subjects found WWMX more "satisfying" than PhC (mean rating of 8 for WWMX versus 6.1 for PhC). In total, only in three columns was WWMX better than PhC in a statistically significant manner. In addition to the "satisfying" rating, the WWMX experience had a higher entertainment value (7.8 versus 6.2). As for the third, "know beforehand" is a subjective measure of how well the subjects thought they knew, before embarking on each task, where to look for the photo. Again, WWMX was rated higher than PhC (7.9 for WWMX, 6.6 for PhC). However, this subjective measure was not backed by the actual task performance of the users, as we demonstrate below. Other than these three measures, there were no significant differences between WWMX and PhC in the subjective task evaluations.

Figure 5 shows results for the experiment's objective measures, both for the Browsing Task (three leftmost columns) and Search Task (two rightmost columns). Each measure is represented by a different column. The corresponding units appear underneath each column, and the corresponding val-

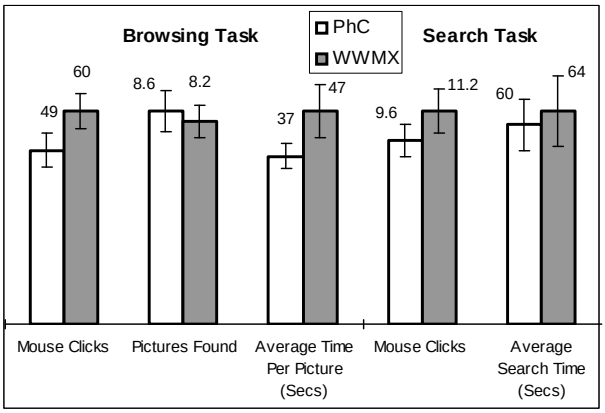

Figure 5: Objective measurements of Browse Task and Search Task (two rightmost columns).

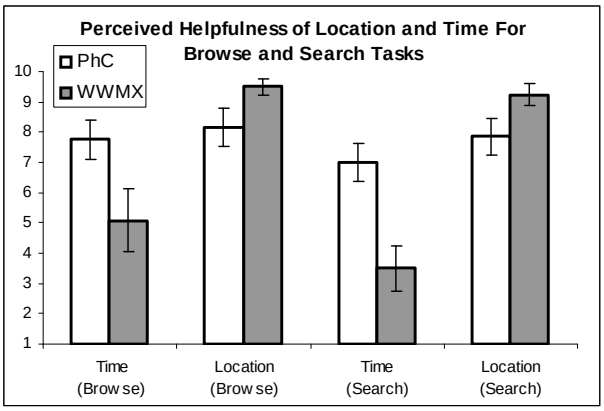

Figure 6: Subjective reports on the use of Location or Time categories in both applications.

ues on top of each bar. We can see that for the Browsing task, the number of photos found by the subjects is statistically equal (8.6 for PhC, 8.2 for WWMX, no significant difference). However, the photos were collected significantly faster using PhC (37 seconds versus 47 seconds). In addition, WWMX required more mouse clicks for the browsing task.

For the Search Task, as shown in Figure 5, there were no significant differences in search time and the number of mouse clicks between the two interfaces.

In addition, we asked the subjects for a subjective evaluation of each application's interface. The subjects rated terms like "complex", "efficient", "helpful", "novel", "intuitive", etc. Out of these measurements, none were significantly different between WWMX and PhC.

Figure 6 compares subjects' subjective sense of whether time and location, respectively, were important for completing tasks on either system. That is, whether the location and time were powerful manipulators in the two applications. The subjects had given a 1–10 rating for the perceived usefulness of each dimension, in each application, for both the Search and Browsing Task. For example, the first column shows users had found the time dimension more useful in PhC than in WWMX for the browse task. Indeed, as seen in the figure, time played a much larger role for PhC during both browsing and searching than it came into play for the WWMX condition. Recall that PhotoCompas attempts to identify and name clusters of photographs that were taken at the same event. In WWMX the time element is instead represented via a visual timeline approach.

The difference in the importance of *location* between the two applications was not quite as pronounced, but still significant. Figure 6 documents this effect in the two location columns for the browse and search tasks. For WWMX, subjects felt that location played a larger role than they did for PhC (ratings of 8.1 and 7.9 for PhC in Browse and Search versus 9.5 and 9.3 for WWMX). Nevertheless, location was clearly important in both interfaces and for both types of tasks.

5. DISCUSSION

Before discussing the results of the previous section, we reiterate that the number of subjects we were able to recruit, given the exacting requirements for subjects' photo collections and time commitment, does not permit broad conclusions. Nevertheless, trends are evident and important to consider for the continuing design of PhotoCompas, WWMX, and other photo browsers.

5.1 Measured and Questionnaire Results

Table 1 more qualitatively summarizes the statistically significant differences between the applications as discovered in the experiment. The applications performed equally for most other metrics and measurements.

Table 1: Summary of statistically significant differences.

Application	*Advantage*
WWMX	Found to be more entertaining and satisfying by participants
WWMX	Participants felt more secure beforehand as to where to find photos
PhC	Browsing task required less time and mouse clicks
PhC	Event/time dimension more useful for browse/search than WWMX
WWMX	Location dimension more useful for browse/search than PhC

The most surprising result of this study is that so little difference emerged in the averaged objective measures of search and browse speed. The WWMX user interface is much more visually oriented than the PhotoCompas interface, which relies predominantly on textual cues and a more analytic mental model. The mouse click count is a particularly suggestive measure, as their function of specifying constraints is the same in both systems. Note that for the browsing task PhC elicited significantly fewer clicks than WWMX. This fact may be one of the contributing factors to PhC's shorter per-picture browse time (Figure 5).

The visual differences were exacerbated by the currently still unrefined screen appearance of PhotoCompas, when compared to the mature WWMX look. The visual difference was noted by virtually all subjects and is reflected partly in the responses to the *entertaining/satisfying* questions, where WWMX shined.

Also surprising was that on average both systems were perceived to be equally *easy*, and both received statistically equal scores on the *confusing* response. We had expected that one system would be perceived to be more straightforward than the other. We had indeed expected that the textual PhC would tend to be more confusing than the much more broadly familiar map-based interface.

These results certainly suggest that PhC's location and event hierarchies were intuitive enough, and together with the automatically generated names and thumbnails, users had a good grasp of PhC's navigation mode. Also, the results may suggest that users utilized the improved time context (the notion of event) in PhC to compensate for the deficiency of the map-less location navigation. Indeed, users have repeatedly asked for the addition of an "events" feature in WWMX as we report below.

We also noticed, through observing the subject during the experiment and subjective feedback, that a while most subjects liked the map-based interface, few subjects implicitly and explicitly expressed their aversion towards the map-based interface in favor of the text-based hierarchical browsing of PhC. This aversion slightly reflected in their performance measures as well. One question that this observation raises is whether there is a strong bipolar trend in people's preference for map-based interface versus text-based interface. The limited size of our subject pool prevents us from answering this question with statistical significance at this point.

5.2 Results from Debriefing Session

We received valuable feedback during the concluding, informal portion of the experiments where we asked participants for open-ended feedback about both systems. Several issues in particular stood out:

- The PhC notion of events was popular, and WWMX was lacking in the Time dimension. Not only participants found the event metaphor intuitive in PhC, they also requested this feature for WWMX as well. Others commented about the handling of time in WWMX, where thumbnails that are presented on the screen (i.e., correspond to the currently displayed map region and timeline) were not sorted by time. Also, the WWMX timeline interaction was not intuitive to some.

- The text-based search mechanisms were too limiting in PhC. subjects requested that keyword search should be made smarter. For example, if a majority of photos had been shot in Yosemite National Park, but one or two were taken in Groveland, a small town near, but outside the Park, then the PhC algorithm would produce the label "Yosemite National Park" for all photos from that area. The term "Groveland" would therefore not occur in the system's label corpus. A search for this term would thus fail. A more intelligent engine would test whether the given location name was within one of the clusters the algorithm had identified, and would return a more helpful answer. In this case, entering of the terms can be aided by auto-completion based on the locations that appear in each user's collection. Such auto-completion is of course also possible for the WWMX interface.

- The size of thumbnails was an issue for many participants, in both interfaces. Several subjects suggested that the PhC thumbnail could and should be smaller. On the other hand, WWMX that displayed smaller images had received the opposite feedback (and so did our work in [8]). The message is that a single thumbnail size truly does not seem enough to span user preferences. Many commercial photo browsers, of course, offer a choice of sizes; the feature is easy to implement given the sizeable resources of desktop machines where scaling operations are fast, and storage is available for precomputed size alternatives.

- Many subjects requested better abbreviated summaries of image clusters. As mentioned earlier, our implementation platform happened to preclude the necessary operations to intelligently choose representative photos from a set of photos. Procedures for choosing images that ensure good summarization of a photo set are still open to research. In other work [7] we showed that choosing images that span the creation time range of the photos in their clusters is effective. The location dimension now offers an additional source of diversity whose coverage in the summary likely leads to good overviews. A summary can, for example, ensure coverage of all locations that are heavily represented in the photo set to be summarized. Image analysis tools can add yet other criteria for compositing effective visual summaries. The summary can, for example, be constructed to include representatives with widely differing color compositions. When face detection is available, summarization algorithms can be made to prefer images that contain faces. Finally, the additional technical information that digital cameras add to the image files they produce is another source of support for selection decisions. For example, cameras include aperture and shutter speed settings in each photo file. This information can be used to ensure that both close-ups and distance shots are included in each summary.

- For both applications, participants asked for additional ways to add and manipulate the metadata. On example is renaming events in PhC to reflect the actual content of the event (e.g. "Grandma's birthday" instead of "San Francisco"). Another example is adding information about the people in the photos.

- A calendar view was often requested for PhC, in addition to the flat event breakdown. Recall that the PhC interface grouped events together by the year, without a month level in between. For some participants, there were too many events in every year, and another level was required. In addition, some participants asked for a "month" or "season" category that is independent of the year: often, they remember an event occurred in a specific month or season, but was not sure about the year. Indeed, as mentioned earlier, we use such a category in our full interface version [13] that was not studied in this paper.

- Alternatively, users asked for further breakdown of the high-level event to lower-level events, especially when the high-level event consisted of a few photo-taking days. This feature is certainly feasible, and we have implemented it in some of our other implementations [7, 8].

Finally, and expectedly, many participants noted that some combination of the two applications would be beneficial. As one participant put it, "these are two metaphors I need at various times".

6. RELATED WORK

A number of approaches to the photo collection management problem exist. First, tools can facilitate manual annotation. For example, some tools allow one annotation to be easily associated with many photographs (e.g. [17]). Even good tools seem unable, however, to remove the reality that annotation is cumbersome and time-consuming for consumers and professionals alike.

Second, methods for fast visual scanning of the images, such as zoom and pan have been developed. These tools (e.g [1]) are quite helpful for viewing several hundreds of photographs efficiently. They do not, however, scale to manage tens of thousands of images. They are also not designed for searching over the collection, as they focus on browsing activity.

Automatic image analysis tools aim to supply the search component. This approach is very difficult, and precise analysis is not yet feasible. The special case of face detection has become reliable enough that it can be deployed in this usually non-critical application of consumer photography. Image analysis in balance, though, is not yet practical for the meaningful and comprehensive organization of photo collections. For a summary of content-based image retrieval systems, see [21].

Finally, many different research projects [3, 5, 12, 15] have addressed the problem of automatically organizing a photo collection based on photo metadata — albeit usually limited to time metadata only, as no one of these papers had experimented with location metadata. Most of the mentioned papers looked at detecting time-based events in the collection, corresponding to the natural way users think about their photos [4, 16].

Introducing location, a number of systems have been developed that deal with presenting geo-referenced data, especially within the Geographic Information Systems (GIS) community [2, 10, 18, 11], but nearly all of them rely solely on a map based interface, and none of the GIS research efforts cater towards browsing and searching in the context of personal photo collections. In [2] the application displays geo-referenced photos as points on a zoomable map interface, but the user is unable to see any of the actual photos until a specific point is selected.

There are several ad-hoc systems that incorporate a map interface for organizing personal photo collections. In [19] the author presents a system called GTWeb, that automatically generates web pages with maps and other annotations from digital photographs and corresponding GPS track data. The points at which the photos were taken are graphically represented as trails on a (non-interactive) map. However, GTWeb focuses on representing photos from long trips, and does not support hierarchical navigation by event or location. The MediaFinder [9] system offers a flexible interface for the user to organize their personal media items spatially in a variety of semantic structures. MediaFinder requires manual specification of the spacial regions and hierarchy, and does not provide much support for non-spatial organizations (such as by time).

7. CONCLUSIONS AND FUTURE WORK

We showed in a controlled experiment that, against intuitive expectation, a textually oriented browser can enable as good a user performance as a map based interface to a time- and location-stamped personal collection of photographs. Subjects performed searching and browsing tasks just as quickly and completely with the textual approach as with the alternative.

To compensate for the deficiency in location-based browsing, our browser was aided by an enhanced time-based support in form of events. However, participants were also able to efficiently navigate the text-based location hierarchy.

However, the visual appeal of the map based approach was not lost on our subjects. Many suggested a combination of the two facilities. Such a combination is indeed natural. A challenge will be to accomplish the fusion such that drawbacks of the map are compensated for. The drawbacks include inefficient screen estate usage and the confusion that portions of the population experience when studying maps. Once that fusion has been prototyped, this study will serve as a baseline for measuring its success.

The result of our study is also important for guiding designs that cannot rely on maps. A prominent example is the access of photo collections on small devices. The limited screen size on those platforms severely handicaps the use of maps as a primary interface element.

Beyond the above mentioned attempt to fuse the approaches we now examined in isolation, a number of explorations remain to be undertaken. The integration of geo-aware keyword search across the synthetically derived place and event names will be an important improvement. A more sophisticated summarization of sub-collections will be needed. Limiting the summarization to the sub-collection's first four images is insufficient; the solutions may be picking more representative photos, or zoom/adaptive thumbnail methods.

Finally, once the location of photo shots is known, a number of secondary information about the shots can be derived. For example, we have in separate implementations [13] added information about weather conditions at the photos' time and place. Such information is maintained in climatological archives. The effectiveness of such additional clues remains to be examined.

Digital photo collections are growing rapidly. Not only are consumers accelerating their image production as inexpensive digital cameras penetrate the market. Professionals, such as biologists who collect images in the field are also in the process of moving to digital imagery. Effective access to the resulting archives will be the foundation for making these researchers successful.

8. ACKNOWLEDGMENTS

We thank Kentaro Toyama and Ron Logan of Microsoft Research, the creators of WWMX, for their generosity in making their tools available to us. Marti Hearst and Kevin Li of UC Berkeley's SIMS made Flamenco available to us and even made code modifications to adapt the program to the needs of our experiment. Andy Kacsmar, our system administrator, bravely suffered through the experiment and its weekend system uptime demands.

9. REFERENCES

[1] B. B. Bederson. Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In

Proceedings of the 14th annual ACM symposium on User interface software and technology, pages 71–80. ACM Press, 2001.

[2] D. Cavens, S. Sheppard, and M. Meitner. Image database extension to arcview: How to find the photograph you want. In *Proceedings of ESRI Users Conference*, 2001.

[3] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. Temporal event clustering for digital photo collections. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 364–373. ACM Press, 2003.

[4] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss. Requirements for photoware. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 2002.

[5] U. Gargi. Consumer media capture: Time-based analysis and event clustering. Technical Report HPL-2003-165, HP Laboratories, August 2003.

[6] Google inc. http://www.google.com.

[7] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, 2002. Available at http://dbpubs.stanford.edu/pub/2002-4.

[8] S. Harada, M. Naaman, Y. J. Song, Q. Wang, and A. Paepcke. Lost in memories: Interacting with large photo collections on pdas. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries*, 2004.

[9] H. Kang and B. Shneiderman. Exploring personal media: A spatial interface supporting user-defined semantic regions. Technical report, 2004.

[10] Y. Leclerc, M. Reddy, L. Iverson, and M. Eriksen. The geoweb - a new paradigm for finding data on the web. 2001.

[11] P. Longley, M. Goodchild, D. Maguire, and D. Rhind. *Geographic Information Systems and Science*. John Wiley & Sons, 2001.

[12] A. Loui and A. E. Savakis. Automatic image event segmentation and quality screening for albuming applications. In *IEEE International Conference on Multimedia and Expo*, 2000.

[13] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatically generating metadata for digital photographs with geographic coordinates. In *Proceedings of the Thirteenth International World-Wide Web Conference*, 2004.

[14] M. Naaman, Y. J. Song, A. Paepcke, and H. G. Molina. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries*, 2004.

[15] J. C. Platt, M. Czerwinski, and B. A. Field. Phototoc: Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research, February 2002.

[16] K. Rodden and K. R. Wood. How do people manage their digital photographs? In *Proceedings of the conference on Human factors in computing systems*, pages 409–416. ACM Press, 2003.

[17] B. Shneiderman and H. Kang. Direct annotation: A drag-and-drop strategy for labeling photos. In *Proceedings of the International Conference on Information Visualization*, May 2000.

[18] T. R. Smith. A digital library for geographically referenced materials. *Computer*, 29(5):54 – 60, MAY 1996.

[19] D. Spinellis. Position-annotated photographs: A geotemporal web. *IEEE Pervasive Computing*, 2(2):72–79, 2003.

[20] K. Toyama, R. Logan, and A. Roseway. Geographic location tags on digital images. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 156–166. ACM Press, 2003.

[21] R. C. Veltkamp and M. Tanase. Content-based image retrieval systems: A survey. Technical Report TR UU-CS-2000-34 (revised version), Department of Computing Science, Utrecht University, October 2002.

[22] W. Wagenaar. My memory: A study of autobiographical memory over six years. *Cognitive psychology*, 18:225–252, 1986.

[23] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the conference on Human factors in computing systems*, pages 401–408. ACM Press, 2003.