

Context Data in Geo-Referenced Digital Photo Collections

Mor Naaman, Susumu Harada, QianYing Wang[†],
Hector Garcia-Molina, Andreas Paepcke
Stanford University

mor, harada, hector, paepcke@cs.stanford.edu, [†]wangqy@stanford.edu

ABSTRACT

Given time and location information about digital photographs we can automatically generate an abundance of related contextual metadata, using off-the-shelf and Web-based data sources. Among these are the local daylight status and weather conditions at the time and place a photo was taken. This metadata has the potential of serving as memory cues and filters when browsing photo collections, especially as these collections grow into the tens of thousands and span dozens of years.

We describe the contextual metadata that we automatically assemble for a photograph, given time and location, as well as a browser interface that utilizes that metadata. We then present the results of a user study and a survey that together expose which categories of contextual metadata are most useful for recalling and finding photographs. We identify among still unavailable metadata categories those that are most promising to develop next.

Categories and Subject Descriptors

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

General Terms

Human Factors

Keywords

geo-referenced digital photos, photo collections, context

1. INTRODUCTION

Managing personal collections of digital photos is an increasingly difficult task. As the rate of digital acquisition rises, storage becomes cheaper, and “snapping” new pictures gets easier, we are inching closer to Vannevar Bush’s 1945 Memex vision [3] of storing a lifetime’s worth of documents and photographs. At the same time, the usefulness

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM’04, October 10–16, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-893-8/04/0010 ...\$5.00.

of the collected photos is in doubt, as the methods of access and retrieval are still limited.

The existing approaches towards the photo collection management problem can be categorized into three main thrusts. First, there are tools to enable and ease annotation [16]. However, annotation is still cumbersome and time-consuming for consumers and professionals alike. Second, methods have been developed for fast visual scanning of the images like zoom and pan operations [2]. These tools may not scale to manage tens of thousands of images. Finally, image-content based tools [18] are not yet, and will not be in the near future, practical for meaningful organization of photo collections. While low-level features can be extracted, the semantic gap between those and recognizing objects (and furthermore, topics) in photographs is still wide.

The good news is that automatically collected metadata has been shown to be helpful in the organization of photo collections. In [6, 7] we demonstrate how the timestamp that digital cameras embed in every photo is effective in the construction of browsers over photo collections that have not been annotated manually.

Beyond time-based automatic organization, technology advances have made it feasible to add location information to digital photographs, namely the exact coordinates where each photo was taken.¹ As location is one of the stronger memory cues when people recall past events [19], location information can be extremely helpful in organizing and presenting personal photo collections.

We have implemented PhotoCompas [13], a photo browser that exploits such geo-referenced photographs. PhotoCompas uses the time and location information to automatically group photos into hierarchies of location and time-based events. The system was proven effective [11, 13] for users browsing their personal collections.

We have now extended PhotoCompas; in addition to employing the time and location metadata to automatically organizing photo collections, the system deploys time and location *also* as generators of context. We extract additional metadata about each photo from various sources. We then integrate the metadata into the browser’s interface. The metadata provides *context* information about each photo.

For example, we obtain weather information about each photo. The time and place where the photo was taken allows us to retrieve archival data from weather stations that are local to the photo’s exposure location. Similarly, given time

¹There are a number of ways to produce “geo-referenced photos” using today’s off-the-shelf products. For a summary, see [17].

Time	2002 (398) 2003 (3306)	Time of Day	Afternoon (12pm-5pm) (1873) Early morning (3am-6am) (22) Evening (5pm-8pm) (650) Late night (12am-3am) (28) Morning (6am-12am) (223) Night (8pm-12am) (908)
Location	Cambodia (151) France (167) Hungary (176) Israel (670) Italy (146) Sri Lanka (512) Thailand (60) United states (1822)	Weather Status	Clear (944) Fog (2) Haze (135) Heavy rain (6) Light rain (237) Light rain showers (3) Light snow showers (3) Mist (61) Mostly cloudy (373) Overcast (110) Partly cloudy (490) Patches of fog (1) more...
Elevation	-2000--1001 (36) -1000--1 (327) 0-999 (4028) 1000-1999 (151) 2000-2999 (53) 3000-3999 (43) 4000-4999 (57) 10000-10999 (85) 11000-11999 (39) 12000-12999 (37) 13000-13999 (40) 14000-14999 (33) more...	Temperature	20-40 (87) 40-50 (972) 60-80 (760) 80-100 (239) Unknown (1646)
Season	Autumn (sep 21st-dec 20th) (1007) Spring (march 21st-june 20th) (953) Summer (june 21st-sep 20th) (1059) Winter (dec 21st-march 20th) (685)	Time Zone	-5 (8) -7 (180) -8 (1634) 1 (499) 2 (670) 5 (512) 7 (211)
Light Status	Dawn (47) Day (2367) Dusk (495) Night (789)		

Figure 1: The metadata categories generated by our system, as shown in the interface opening screen.



Figure 2: A subset of the "Sri Lanka dusk photos" from the first author's collection, detected using contextual metadata.

and location we automatically obtain the time of sunrise and sunset when the photo was taken. PhotoCompas integrates this contextual information in its user interface (see Figure 1). In Section 2 we briefly describe the metadata we thus assemble for each photograph.

While the contextual metadata can serve well as memory cues, it can also imply the content of the image. For example, by clicking on the *Dusk* entry in the *Light Status* section of Figure 1, a photographer requests to see only photos that were taken when dusk had fallen at the exposure location. Figure 2 shows the result of this action, further restricted to show only photos from Sri Lanka.

Our focus here is on (i) measuring how effective users believe our particular set of contextual metadata to be for photo retrieval, (ii) observing which of this metadata they actually take advantage of when searching through our interface, and (iii) exploring what other contextual metadata would be profitable to capture in the future.

To this end we gathered and analyzed several data sets by means of a user study and a separate survey. In the user study we had subjects find photographs from their own collection by interacting with our metadata enriched PhotoCompas browser. We recorded their paths through the interface. Also in the user study we asked each participant to grade 24 potential categories of contextual information by

how useful they believed the additional information would be for helping with retrieval. In the survey, which involved a larger number of participants, we had participants describe some of their own photos from memory. We also asked them to grade the same 24 metadata categories in terms of how well they were remembered. We then compared the collected data and conclude which of the 24 categories seems to be effective for photo collections.

The practical results of this work are recommendations on which metadata sources would be particularly profitable to create or tap into in order to make photo collections more easily accessible.

The methods we describe in this paper may extend beyond photo collections to assist in the management and retrieval from, for example, an all-encompassing personal data store [5]. Alternatively, some of the metadata may be useful for annotation and retrieval from global repositories, such as a set of geo- and time-referenced news reports.

The rest of this paper is organized as follows. In Section 2, we describe the contextual metadata that we currently extract from various available sources, given time and location information for each photo in the collection. Section 3 discusses the use of these metadata in the user interface. Section 4 describes the user study we devised to assess the usefulness of our added metadata and presents the results of the study. In Section 5 we describe the methods and results of the survey, and how they relate to the user study's results. We conclude in Section 6.

2. METADATA CATEGORIES

The metadata categories our system generates and utilizes in the interface are listed below.

2.1 Location

The location recorded for each photo refers to the location of the *camera* at the moment of capture. Another interesting location is the location of the *subject* of the photo, which we assume is approximately similar to the camera location.

The location is represented by latitude and longitude, and must be translated to a useful location representation. A map is one option [17], though we have shown that browsing a textual location hierarchy can augment map-based location browsing, or even replace the map [11] (for example, when screen real-estate is limited).

The text-based location is based on an off-the-shelf geographical dataset that enables querying with a (*latitude*, *longitude*) pair to get the containing country, province/state, county, city (if any) and park (if any). Additional location context can be supplied by gazetteers such as Alexandria [8], which can be queried for nearby landmarks to each photo.

In our PhotoCompas implementation, we group the photos into geographic *clusters* that are expected to make sense to the user, but do not necessarily correspond to the administrative division of provinces, counties, cities, parks etc. For example, a user that lives in San Francisco may have clusters for the San Francisco area, Yosemite area and the East Coast; the details of how we generate the clusters and assign them with meaningful geographic names (based on the geographical dataset) appear in [13].

2.2 Time of Day

Knowledge about the time of day a photo was shot can be used when searching for an image in a collection (e.g.,

one remembers a certain photo was taken late at night). In most cases, however, users set their camera clock once, usually when they operate the camera for the first time. The clock is used to timestamp all the photos taken with the camera. Indeed, this one-time setting is sufficient to sort all the pictures by time, or to associate pictures with the (approximate) month and date in which they were taken. The pitfall appears when the user travels to a different time zone. Most of the users do not bother re-setting their camera's clock, resulting in timestamps that inaccurately portray the time the picture was taken. For example, a photo taken in Israel in the morning, using a camera set to Pacific Standard Time, would appear to have been taken at night because of the 10-hour difference.

This problem can be solved given the location information where each photo was taken, and the original time zone according to which the camera clock was set. This information is sufficient to compute the local time for each photo. In the future, location-aware cameras could automatically detect the local time, as do cellular phones today.

In the meantime, we compute the local time using a geographical dataset containing the world's time zones. The dataset can be queried by the photo coordinates, returning the correct time zone for the photo. The next step is simply calculating the offset between the original and actual time zones and applying it to the photo time to get a local time.

2.3 Light Status

People's perception of the time is sometimes not derived from the clock time, but rather from the daylight status. For example, people may recall a certain picture was taken when it was dark outside; around sunset; before sunrise; etc.

Given the local time and location for each photo, one can find how many minutes away from the sunset and sunrise each picture was taken. One way to do it is implement an algorithm, and compute the sunset and sunrise time for each photo location and time.

Instead, we use the US Naval Observatory Web service². The service returns a complete set of daily sunset and sunrise times for a (*year, latitude, longitude*) query. To give an idea about the number of necessary queries, only 42 network requests were required (accurate to integer lat/long values) for a personal collection of 3706 photos spanning 13 months and 8 countries, as the results can be cached and reused for the each (*year, latitude, longitude*) value in the collection.

As people are not expected to remember exactly how much time before or after sunset or sunrise their pictures were taken, we group photos into *day, dusk, night* and *dawn*. In our current implementation, the dusk grouping includes all photos taken within one hour before or after sunset; the night photos include all photos taken one hour after sunset to one hour before sunrise, and so on. This arbitrary grouping may be a somewhat limiting. For example, northern locations have longer sunsets. We are considering ways to overcome this problem.

2.4 Weather Status and Temperature

Often, people can filter photos using weather information: they recall a certain event occurred on a stormy night (even if the event took place indoors), another event on a clear day, etc. In addition, people may remember the outside temperature at the time the picture was taken ("it was freezing!").

²<http://aa.usno.navy.mil/>

We use the Weather Underground³ Web service to get weather information. Historic data can be queried by a (*zip-code, date*) pair or a (*weather station, date*) pair for weather outside the United States; geographic datasets that allow us to translate any (*latitude, longitude*) pair into a zip code or weather station are commonly available. The results of a query to the server can be cached and used for all photos taken on the same day and in the same area, reducing the required number of queries. Again, the number of required queries is small in comparison to the number of photos. For 1823 US photos, for example, spanning 13 months and 70 different zip codes, only 139 queries to the Web service were required when caching the results.

The weather data we get for each day is an hourly (approximately) report of the weather conditions (e.g., "rainy", "clear") and temperature. We annotate each photo with all weather conditions that appear between two hours before and after the photo (a single photo can be associated both with "rainy" and "cloudy", for example). The temperature is computed as the average of temperatures measured in the hours around the photo time. We used actual measured temperatures; another option is using perceived temperatures, e.g., factoring in the windchill.

2.5 Events

In personal photo collections there is a strong notion of events [4, 14]. People often think of their photos in terms of events, and usually associate each photo with a certain event: a wedding; a vacation; a birthday party etc. The "event" can therefore be considered useful context.

Computer systems today cannot detect the essence of an event (e.g., "birthday"). However, the technology is mature enough to be able to automatically group photos that were taken during the same event [4, 6], especially when location data as available, as we show in [13]. The get the essence of the event, one might consider correlating events our algorithms identify with the photographer's calendar. If the event is public (e.g., a wedding, a baseball game, or a music concert), then sharing data between users may help in automatically or semi-automatically assigning a meaningful name to the event (based on systems such as [12, 15]).

In [13] we describe how PhotoCompas automatically detects events using the time and location information in each photo. The events are annotated with a place name (based on the location of the photos in the event) and time to supply the full automatically-derived context, e.g. "San Francisco, 7th July 2004".

2.6 Additional Categories

Other metadata categories that we derive from the location- and time-stamps of the photographs are: elevation — available either from the GPS data or from a digital elevation model; season (autumn, winter, spring, summer) — by the date in which the photo was taken (of course, this will differ between hemispheres). We also show the time zone (offset from GMT) as a separate category in addition to using it to compute local time.

Many other categories could be produced for browsing photos using current image processing technology, although we have not, so far, integrated such methods into our work. For example, starting with the number of people that appear in the photo (detection technology is quite mature [9] while

³<http://www.wunderground.com/>

recognition is not), whether it was taken indoor or outdoor [10], prominent colors or other content data [18], etc. In this paper we emphasize contextual metadata that pertains to location and time only, but we included many additional categories in our survey as described in Section 5.

3. INTERFACE

While our work is deploying the contextual information explicitly in a user interface, it must be noted that the context can also be used implicitly, e.g. for determining similarity between photos, ranking search results and so forth. Focusing on the interface, we describe the prototype interface we used to initially test the contextual metadata we produced. Before we get to that, we describe the requirements from an “ideal” interface, given this contextual information.

An interface that will enable effective usage of context must be:

- Non-intrusive. The context selection/filtering mechanism should not take too much screen space, nor create clutter that requires additional mental effort.
- Simple and clear. For example, weather conditions should be represented by very few general groups (sun, clouds, rain, snow, wind) and not by dozens of possible weather conditions (showers, thunderstorms and so forth).
- Able to accept range specifications. For example, being able to specify “it was cold — between 0 and 25 degrees” for temperature; specify a calendaric range; a time-of-day range (“sometime between 6am and 8am”).
- Allow exclusion of groups of photos easily. E.g., “any time but night”; “it wasn’t raining”.
- Be flexible enough to simultaneously apply multiple filters (“it was a cold, rainy morning”).

We used the Flamenco HTML-based toolkit [20] to create the interface for our PhotoCompass, and in particular, to enable browsing of the contextual metadata. The authors of Flamenco kindly made their implementation of a general metadata browser available to us. That system does not meet all the requirement laid out above. However, Flamenco 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 the contextual information is indeed useful, and have more information pertaining to how people use, or would like to use, this context.

In every step of the interaction with the Flamenco interface, the user can select from a number of values under each category. For example, “United States” and “Sri Lanka” appear under the Location category in Figure 1. After a user clicks on a value, the interface shows only the photos that match that value, as well as other category values that were chosen earlier (e.g., “dawn”). Further refinement is then possible using both other categories and more detailed groupings of the chosen category (e.g., locations within the given country).

4. USER STUDY

We set out to test the use of the contextual metadata on real-life, large personal collections of geo-referenced digital photos. Our goals were to discover:

- Whether participants are able to use the contextual information to browse for photos.

- Which of the context metadata that can be derived using location and time are useful.
- Whether the results correspond to the outcome of the survey. This is discussed in Section 5.

We first describe the user study setup, and then the methods and results.

4.1 Statistics and Setup

It is practically impossible to find subjects with personal geo-referenced photo collections, as the geocoded photo technology is not widely available to consumers yet. Instead, we used a “location stamping” tool. The Location Stamper⁴ allows users to retrospectively mark their photos with a location, by dragging and dropping their photos onto a map. Still, recruiting enough suitable users was a difficult task, as we had also experienced in previous photo-related studies [7, 11]. Some of the challenges included the requirement of a large enough collection and a considerable time investment. At the end of the subject search process, we were able to recruit nine subjects for our user study.

The participants’ ages ranged from 18 to 32, with the highest representation in the 20s. Six subjects were male, and three were female. Seven of the photo collections exceeded 1,000 photographs, and the average collection size was 1,876 images. The average time span of the collections was 22 months. On average, each collection had photos from four countries, and 82% of the photos were taken in the United States.

4.2 Method and Results

The study was executed using the subjects’ own personal collection of photographs. Each participant completed one task. In this task, we asked the participants at the beginning of the session to mentally recall three of their photographs. Then, we asked them to navigate to each of these photos in turn, applying constraints by manipulating the Flamenco user interface, as described in Section 3. For the user study, we removed the location context (country, place names as describe in Section 2.1) from the interface. We also removed the event breakdown (Section 2.5). The reason for the removal is that we felt the direct location/time context information would dominate over the others, especially as the collections span only a few years (i.e., not many photos from each location). Participants were thus faced with six available contextual categories: temperature, weather status, elevation, time of day, light status and season.

We logged the order in which subjects clicked on the different metadata categories in every trial. We focused on the first few clicks in each trial. Beyond that, the data was too spotty (often, the choice of category to click on was very limited after the first 2–3 clicks).

In addition, we used a questionnaire, parts of which were handed to the participants after they completed each search task, and another part upon completion of their session.

We first report on the actual category ‘clicks’ the participants applied in the different trials, looking for the different photos. Then we discuss the questionnaire.

The click count is summarized in Figure 3, showing how many times each category was clicked in the first or second click of each trial. By far, most user trials (12 out of a total of 27 trials) started by clicking on the season corresponding

⁴Available from <http://wmx.org>

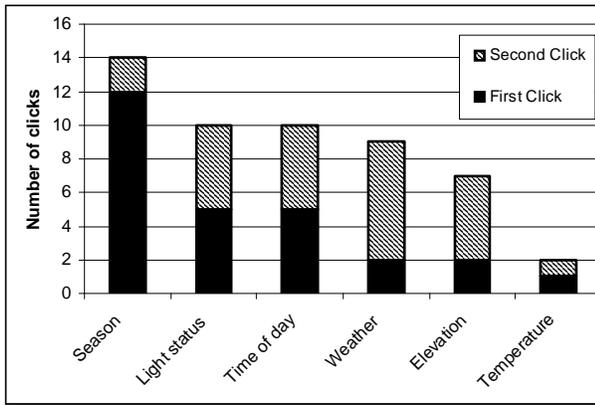


Figure 3: Usage of the different metadata categories within the first two clicks in each trial.

to the photograph they were looking for during that trial. The next most useful categories were light status and time of day, which we suspect were used interchangeably by the participants. Each of these categories was used within the first two clicks in 10 out of 27 trials. The weather conditions category proved useful as well, as it was clicked within the first two clicks 9 out of 27 times. More notably, participants preferred the weather conditions category to the temperature category; the latter was only used twice within the first two clicks.

In the first part of the questionnaire, we asked the participants after each trial to rank the categories that were most used by them during that specific trial. The results are consistent with the order of the clicks. Indeed, participants attested that time of day, season and light status (in this order) were effectively used much more frequently than weather, temperature and elevation.

In the second part of the questionnaire, the participants were asked to freely specify other contextual metadata categories that would have helped them look for that trial’s photo. Prominent answers to this open-ended question were number (or presence) of people in the photo (five participants listed this category, for 13 different photos); indoor/outdoor classification (5, 7); type of activity (2, 2); and prominent colors (2, 2). This part, and its results, correspond to the independent recall part of the survey (see Section 5.2).

At the end of all trials, each participant ranked a group of 24 contextual metadata categories on how useful they would have been in retrieving each photo in the trials. We used a 10 point Likert scale. For example, subjects were asked to rank how useful the data regarding number of people in the image would have been when looking for the specific image. This part of the questionnaire corresponds to the assisted recall rating part of the survey, where we asked the participants how well they *remember* the specific properties of the photo, instead of rating the *usefulness* of the data as we do here (see Section 5). We leave the detailed discussion to Section 5. It is worth mentioning that within this part of the questionnaire, users ranked the metadata that were available in the interface in a consistent manner to the measured click order and the first part of the questionnaire.

5. SURVEY

Independent of the formal user study that we described in the previous section, we administered a survey to 35 participants. This survey allowed us to gather information about the usefulness of various contextual metadata from a broader number of people than was practical in the context of the user study. None of the experiment’s subjects were also in the set of participants we surveyed. In particular, we sought in the survey to understand which metadata categories are useful for *mentally recalling* photographs, rather than for *finding* a photo in a collection, as had been the focus in the user study’s questionnaire.

The survey consisted of two parts that respondents worked on in order. In the first part of the survey, the *independent recall*, we asked the participants to mentally recall any one to three of their personal photographs. Once they had a photo in mind, the survey had them describe the photo in prose in as much detail as possible, and with as much context about the surrounding environment and conditions as they could remember. These expositions usually amounted to about half a page of handwritten text. Through this method we harvested a total of 83 descriptions from the 35 participants.

The open-ended nature of the survey’s photo descriptions provided an opportunity to discover the range of metadata that is at work in mental recall, independent of the facilities we provided in our computer implementation.

To this end we coded all the prose descriptions, extracting which kind of metadata about the respective photo was being used in each sentence of the descriptions. We then ranked these metadata categories by the number of times that respondents made use of them in their descriptions.

In the second half of the survey, the *assisted recall*, we asked participants to rate how well they remembered different categories of contextual metadata. These were the same categories that we had asked about in the user study questionnaire and included categories like “time of day”, “indoor/outdoor” and “Other people at the same event (but not in the photo)”.

5.1 Method and Results

We administered the survey in informal settings; participants’ collections were not nearby. Nor had we prompted participants to interact with their collections ahead of time. No computer interaction was involved. The responses were put down on paper. The survey also did not limit participants to digital photographs. Of the 8 photos, 39 were described by female respondents, 42 by men. One respondent was below 18 years of age. Fifteen were in the age range of 18-25, eight were 26-35, two were 36-46, and nine were 46-55 years of age.

We begin with the results of the assisted recall section of the survey, where we asked participants to rate how well they remembered different cues for each photo.

Figure 4 shows a radar graph that summarizes how subjects rated the 24 metadata categories. The radial axis corresponds to the Likert scale ratings. The categories are arranged around the circle. The further out a point is situated from the center of the circle the more highly our respondents rated the corresponding contextual metadata category.

Most of these categories are clearly identified by their labels. Some are more difficult to decipher. These are:

- **Outindr** refers to whether the photo was taken outdoor or indoor.

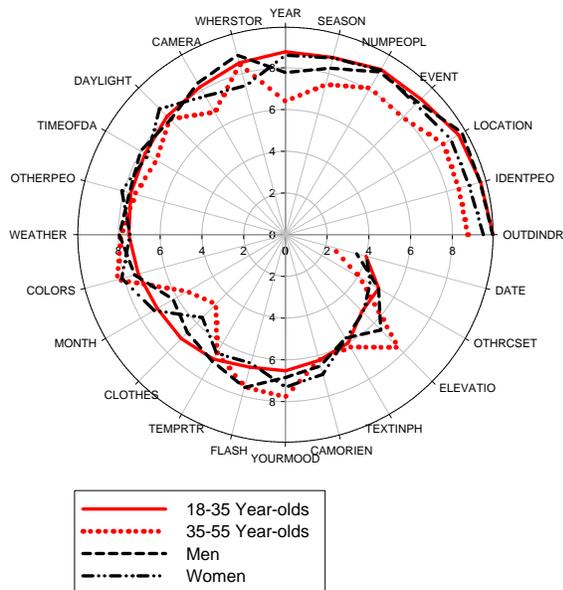


Figure 4: Average values for how well different subject groups remembered different categories for each of their photos.

- **Identpeopl** and **Numpeopl** evaluate the importance of memories about who and how many persons are captured in a photograph.
- **Daylight** is the day/night/sunrise/sunset categorization as described in Section 2.
- **Wherestor** is short for ‘where the photo is stored.’ Example values are “a shoebox”; “my laptop’s drive.”
- **Camera** which camera the photographer had used.
- **Otherpeo** refers to other people who were present at the time, but were not captured in the photo.
- **Yourmood**, of course, refers to the photographer’s mood at the time he snapped the picture.
- **Textinph** refers to text that is visible in the photo, such as a road sign, or the entrance sign to a park.
- **Othercset** is ‘other camera settings,’ or camera settings that are not accounted for separately in the survey.

We had collected the categories from prior pilot interviews. Indeed, the categories in Figure 4 cover all but one of the metadata categories we observed in the independent recall part; see Section 5.2.

For an overview, we show in Figure 4 results for four sub-groupings of respondents. Two graphs shows memory cue ratings by respondents in different age groups; the other two demonstrate differences between genders.

We compared the differences among cue ratings for the consolidated means over all groups in the graph. The results roughly partition the cues into four groups: *outdoor/indoor* stands in a group of its own. It is statistically higher ranked than any of the other cues. A second tier is formed by the cues *number of people*, *identity of people*, and *location*. Moving further counter-clockwise around the graph, *event* through *mood* form a roughly equivalent group. Importance of cues beyond *mood* drops off quickly.

The low values for *date* for all groups confirms the findings of [19]. That study of recalling events also explored the relative importance of different memory cues and concluded that the most important cues are ‘who,’ ‘where,’ and ‘when,’ in that order. However, notice that finer-grained distinctions need to be made than a simple ‘when.’ The *time of day* ranks higher in our data than the more general *date*, and is also a stronger cue than the *month*. In these results at least, the time of day draws even with the memory of which year a photo was taken. This finding points to the necessity of ensuring accurate capture of the *local time* when a photo is shot, as described in Section 2.

After *indoor/outdoor*, The four cues that involve people and location are on average rated highest by the respondents. This result again confirms [19]. The result also reinforces the opportunity that the georeferencing of photographs presents for the automation of organizing photo collections. Also, determining the number of people in a photo is within reach of today’s image analysis technology. Our results clearly call for the integration of this cue into photo browsers. Similarly, the data ranks the indoor/outdoor distinction as important enough to warrant research investment and/or implementation in photo browser applications.

Looking at the graph as a whole we see that most findings are consistent across age groups and gender. However, notice that in most cases the oldest group of respondents ranked *all* memory cues lower than their cohorts, even if relative tendencies match. This observation is quite noticeable for the *year* and *clothes* cues. We can only speculate about the reasons for this finding. The difference in the *year* ranking might be explained by the fact that older respondents have lived through more years, thereby losing this cue as a greatly distinguishing feature.

A last observation on Figure 4 is that the data does not confirm an easily assumed stereotype. One might be tempted to predict that mood during the time of the photo session would be more important a cue for women than for men. The data of this study, at least, suggests that this distinction is not prevalent.

5.2 Independent Recall Descriptions

We analyzed the textual descriptions of photographs, given by participants in the independent recall part of the survey, as described above. Our goals were to see which are the prominent categories when participants freely recall photographs; and whether the findings match the assisted recall findings. The latter question is discussed in Section 5.3.

The dashed (mostly innermost) line in Figure 5 corresponds to the number of times each category was mentioned in the text description of a photo (scaled to fit the 1–10 range). For example, the figure shows that the most popular categories that were mentioned in the descriptions were indoor/outdoor (69 out of 83 photo descriptions touched on this category), the number and identity of people (64 and 56 descriptions) and the location (54). We also checked the *order* in which the categories were mentioned. These top categories were also, by far, the most frequently mentioned within the first three categories for each photo. Overwhelmingly, the identity of people in the photo was the first category to be mentioned (40 out of 83 photos). We do not discuss the order in detail here for lack of space.

The daylight, time of day, and weather, some of the categories we automatically include in our browser prototype,

played an important role in the descriptions (25, 22 and 17 mentions, respectively). The weather descriptions were sometimes quite specific: “really nice, sunny, blue sky, not much of a wind, and it was warm, but I wouldn’t say it was hot.” Another category we used, season, was mentioned 16 times — almost as much as the year (22) and a lot more than the month (9) when the photo was taken. Descriptions that fall into the event category appeared 11 times.

Similarly, emotions that are associated with photos populated the photo descriptions (14 appearances). For example, one respondent described the image of cows tightly packed and readied for slaughter as “unbelievably terrible.” On the other end of the spectrum an excited respondent described the photo of her very young son with “he was standing!” Clearly, the emotional state of a photographer is inaccessible to the camera. Nevertheless, this evidence of strong emotional associations with photographs does suggest features for photo browsers such as easy methods of associating emotions with photos in browsers.

The detailed analysis of descriptions did, however, suggest one metadata facet that is accessible to image analytic methods today: color. The descriptions contained numerous mentions of this facet: “bluish-green ocean,” “mostly black... also subtle shade of golden yellow,” “a nice green spot [of grass where a photo’s subject was standing].” Integration of color-based querying systems [1] in photo browsers may be useful. Of course, reliable semantic annotation based on image analysis is not trivial. For a summary of content-based retrieval technology, see [18].

Finally, we looked at whether the independent recall resulted in cue categories that did not appear in the set of categories we assembled for the assisted recall part. The only category that we had not included in our set, and that was often evoked in the photo descriptions were *objects* (27 appearances). Examples were “a camping van,” “stuffed animals,” “a mansion,” and “the world’s largest lobster.” Clearly, this category is very difficult to detect automatically with today’s technology.

5.3 Recalled Cues vs. Useful Cues

Remember, we presented the subjects of our user study with the same categories of metadata as the assisted recall part of the survey. In the study’s context, however, subjects rated the cues by how *useful* they had been for finding the photo they had chosen to search for on the computer. In the context of the survey, participants had evaluated the cues for how well they *remembered* them for a photo. Is there a difference between how the cues rank for (i) remembering and mentally visualizing a photo, (ii) usefulness for computer-assisted search, and (iii) describing photos in the open ended independent recall part of the survey? Figure 5 provides answers to these questions.

The solid line shows how survey participants ranked the 24 cues for how well they are remembered for each photograph. This line is the aggregate of Figure 4. The dotted line plots the rankings that user study participants assigned to the cues (how useful each cue was in searching). The dashed line, finally, traces the number of times that survey subjects mentioned the respective cue in the photo descriptions of their survey form. These numbers are scaled to match the 0–10 dimension of the graph. The quickly diminishing appearance of this line arises since some cues were never used by survey participants when describing their photographs.

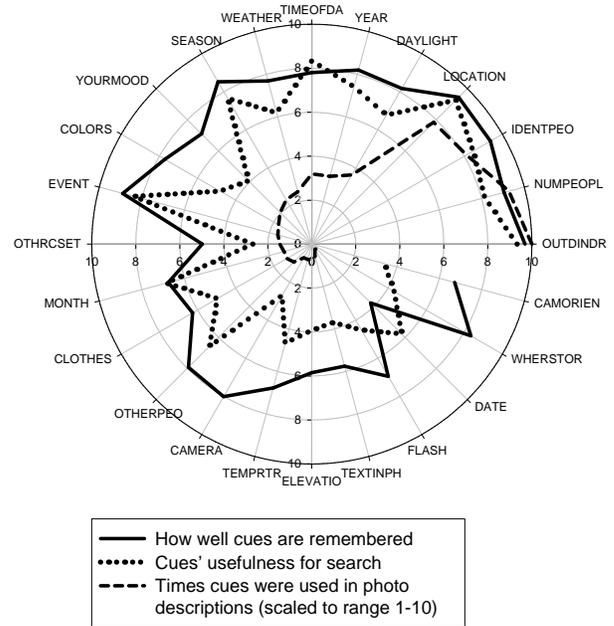


Figure 5: Comparing Perceived Importance of Search Cues, Recall Cues, and Cues Used in Photo Descriptions.

For example, look at the *camera* category. The ‘camera used’ was well remembered (solid line), but not perceived as useful (dotted line), and was mentioned very rarely in the textual descriptions (dashed line).

As Figure 4 shows, outdoor/indoor, location, and the identity and number of subjects in a photo are clearly top choices both for usefulness in searching and remembering; accordingly, they are the top categories that were mentioned in the descriptions. The day/night status when the photo was taken is close behind.

A few noteworthy differences between subjects’ evaluation of cue importance for searching vs. remembering are *where stored*, *camera* and *colors*. According to the figure, these cues are much better remembered than they are useful for search (also, they were mentioned very few times in the open photo descriptions). A simple explanation for the difference in the former two cues is that they do not filter much. For example, most photos were taken by one camera, so it is well remembered but not very useful. On the other hand, *colors* may be perceived as un-useful for search since the users could not imagine the right user interface or language to do so.

Conversely, there is only one category that participants in the study significantly rated higher (as useful for search) than the survey’s participants (as well-remembered). That category is *date*. As discussed above, indeed, people do not remember dates or exact time periods well. However, given that they remember the date, it is useful: it is very easy to find the photo as all photos are timestamped.

6. CONCLUSIONS

The user study and survey we conducted suggest that contextual cues are useful for browsing and managing personal collections of photos.

In particular, location cues prove to be as important as cues involving people in the images. Context involving location and people was rated high in both usefulness for retrieval and being remembered well. In addition, these categories were often listed independently by participants describing their photographs.

Location- and time-derived context such as weather, local time, daylight status and season, proved to be useful as well. Participants interacted with these categories comfortably in a user interface, used them in photo descriptions, and found them useful for retrieval and well-remembered. In particular, local time and daylight status seem to be stronger cues than the date/time of the photo; as collections span many years, they may even be more important for retrieval than the actual year when the photo was taken.

Finally, the outdoor/indoor classification seems not only well remembered and useful, but also on people's minds (as participants mentioned it most in the photo descriptions). In addition, people also seem to remember photo colors, but are not clear if it is useful. Possibly, a simple interface for querying by colors will be beneficial.

This study will serve as a guide towards our, and hopefully others', future effort in automatically integrating contextual metadata into photo browsers. Our ranking of memory and recall values of cues can help decide where research efforts should most profitably be invested: we need to move along the graph of Figure 4 in order to make photo collections increasingly accessible.

7. ACKNOWLEDGMENTS

The authors would like to thank Alan Steremberg and Lawrence G. Pawson from Wunderground.com for giving us access to the weather data for the entire planet.

8. REFERENCES

- [1] J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic. The query by image content (QBIC) system. In *Proceedings of the International Conference on Management of Data (SIGMOD)*. ACM Press, 1995.
- [2] 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.
- [3] V. Bush. As we may think. *The Atlantic Monthly*, July 1945.
- [4] 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*. ACM Press, 2003.
- [5] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. MyLifeBits: fulfilling the Memex vision. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 235–238. ACM Press, 2002.
- [6] 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>.
- [7] 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.
- [8] L. L. Hill, J. Frew, and Q. Zheng. Geographic names - the implementation of a gazetteer in a georeferenced digital library. *CNRI D-Lib Magazine*, January 1999.
- [9] E. Hjelmas and B. K. Low. Face detection: a survey. *Computer Vision and Image Understanding*, 83(3):236–74, SEP 2001.
- [10] J. Luo and A. Savakis. Indoor vs. outdoor classification of consumer photographs. In *Proceedings of the international conference on Image Processing (ICIP)*, 2001.
- [11] M. Naaman, S. Harada, Q. Wang, and A. Paepcke. Adventures in space and time: Browsing personal collections of geo-referenced digital photographs. Technical report, Stanford University, April 2004. Submitted for Publication.
- [12] M. Naaman, A. Paepcke, and H. Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. In *10th International Conference on Cooperative Information Systems (CoopIS)*, 2003.
- [13] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries*, 2004.
- [14] 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.
- [15] R. Sarvas, E. Herrarte, A. Wilhelm, and M. Davis. Metadata creation system for mobile images. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 36–48. ACM Press, 2004.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] W. Wagenaar. My memory: A study of autobiographical memory over six years. *Cognitive psychology*, 18:225–252, 1986.
- [20] 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.