The 32 Days Of Christmas: Understanding Temporal Intent in Image Search Queries

Frank Bentley, Joseph 'Jofish' Kaye, David A. Shamma, John Alexis Guerra-Gomez
Yahoo
San Francisco, CA USA
[fbentley, jofish, jguerra]@yahoo-inc.com, aymans@acm.org

ABSTRACT
Temporal terms, such as 'winter', 'Christmas', or 'January' are often used in search queries for personal images. But how do people's memories and perceptions of time match with the actual dates when their images were captured? We compared the temporal terms that 74 Flickr users used to search their own photo collections, and compared them to the date captured data in the target image. We also conducted a larger study across several billion images, comparing user-applied tags for holidays and seasons to the dates the images were captured. We demonstrate that various query terms and tags can be in conflict with the actual dates photos were taken for specific types of temporal terms up to 40% of the time. We will conclude by highlighting implications for search systems where users are querying for personal content by date.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Photos; Search; Date; Temporal; Queries, Memory

INTRODUCTION
Along with the growth of the smartphone, personal photo collections have grown in size as it has become easier to capture large numbers of photos. These photos are now increasingly being stored in collections on computers or in the cloud, with 1.8 billion photos being uploaded per day [6]. But how can people find the one photo they are looking for out of a collection of many thousands of personal photographs?

It seems obvious, but the difficulty in photo search is inherently one of description. That is to say, photos are images, but we are using words to find them. If you have a document, the words which you use to search are usually in the document that you are trying to find. But there is another layer of disconnect in image search. Recent advances in machine vision begin to describe the objects in the photograph: 'cat', 'tree'. However some of the most salient facts that people remember about photographs include the context around them. Specifically the time, date, location, and sharing history of a photo have been shown to be salient aspects that people remember [2, 7, 8, 10].

But how accurately do people remember these details? When trying to query photos from their collection, how accurately would people remember aspects of the date (year, month, season, etc.) or location of the event? This is compounded by the problem that everyday perceptions of holidays and seasons are much broader than strict definitions of those events. For example, a company Christmas party may happen any time between late November and early January. Likewise, if it is cold and snowing on the 3rd of December, it feels like winter, even though "real" winter is still almost a month away, beginning on the 21st of December.

In this note, we will explore user generated queries on 1,492 personal images from 74 diverse participants on a large photo sharing site (Flickr) to better understand how people think about dates in their own photo collections, as well as the details that people remember or misremember about their past events when searching for photographs in their collections. We will also explore a large-scale dataset of tags from Flickr images and relate temporally-based tags to the dates when these photos were captured. Tying both datasets together, we will conclude with implications from this work on the creation of tools for photo organization and search.

RELATED WORK
The CHI community has been studying photo organization for some time, including ways to query personal photo collections based on metadata in the photos. In the mid-2000s, as camera phones were beginning to enter the mainstream, many researchers explored how we could automatically tag photos with relevant mobile-sensed contextual information, such as temporal, spatial, and social metadata [4]. In particular, time and location could be used to infer data such as daylight or weather conditions. Other work [7] showed that participants were fairly bad at remembering the exact date of a photo. However, specifics as to which aspects of the date participants got wrong, as well as how far off they were in their errors was not explored in this work, making it difficult to design the details of temporal search systems from the data presented. Others

CHI'16, May 07–12, 2016, San Jose, CA, USA.

DOI http://dx.doi.org/10.1145/2858036.2858255.

examined how people remember a photo using a rigid interface to direct participants down traceable paths (e.g. Year→Season or Year→Month) to find photos in their collections [2]. While time, location, and people in the photo were the most salient aspects of a photo, dates themselves were often misremembered and often off by one major unit such as a month or a season in this small-scale lab study. In this work, we wanted to scale up this task, with 74 participants and years worth of photos.

Searching for photos is important, as many researchers have shown people rarely manually tagging their photos so that they could be found again. In their Requirements for Photoware work, Frohlich et al. [5] explored audio annotations and other mechanisms for adding additional narrative information to photos. Ames et al. explored Requirements for Mobile Photoware [1] where annotating images is a large part of the focus. Overall, this work highlights the need for tools to better help people find photos in their growing collections. As mobile camera phones continue to be a part of our daily lives, our collections will continue to grow, and recalling older images becomes more difficult. We set out to explore what people remember about the temporal aspects of their images, some many years old, in our study.

METHODS

We studied this question through a mixed-methods approach incorporating both user-generated queries for specific photos by 74 diverse participants, and a second study of the metadata of large numbers of photographs on Flickr. Together, these studies explored people's temporal views of particular photos and events, and led us to specific ways to modify search algorithms to move beyond rigid definitions of temporal terms to meet the likely intent of the searcher.

Interactive Survey

In early 2015, we conducted an online study with 74 active Flickr users (with at least 500 photos) from the United States. 53% of participants were female; 60% were aged 26–35 and 36% were aged 36–45.

The study gathered queries that participants entered to find their own images. We first required participants to log into their personal Flickr accounts and then selected 20 photos from their collection using the Flickr API. To achieve diversity in the images, and automatically select images that a person might actually want to search for, five of the photos were from the set of 100 most recently taken photos, five had people in them (as identified by the Flickr recognition engine), five were taken outdoors, and five came from a set of high Flickr "interesting" scores (often with many favorites and comments). If participants did not have five images in a category, the remaining images were chosen randomly from the remainder of their collection. From the initial set of 20, participants were able to choose images that they did not want to share with the research team, and these images were replaced with additional images from their collection before proceeding.

For each image, participants were asked to provide a query that they would use to find this image in their Flickr collection. They were given a free text field, similar to the existing Flickr search bar, and nothing in the survey through this point mentioned any particular interest in times, location, or other ways to find photos. After providing a query for all 20 images, additional questions were asked on another page of the survey to probe on specific aspects of the images. The survey concluded with basic demographic information. On average, the survey took 15 minutes to complete, and participants were paid $4.

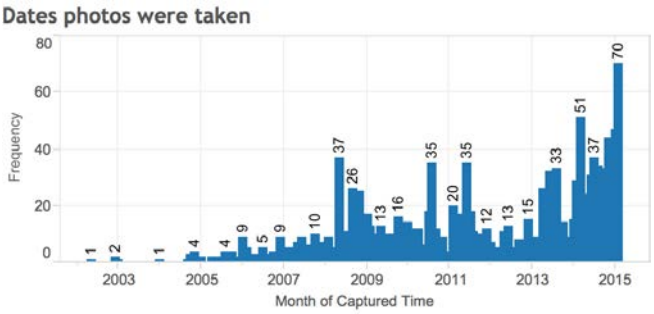

Figure 1. The dates that the photos in our survey were taken. Note that the median date is from 2011, and there are many older images, making the task of searching by date more difficult, yet realistic as photo collections grow.

Data Footprints

To explore the patterns we saw in our $n = 74$ sample at larger scale, we queried the full Flickr database to extract the image-taken date field from each photo that the owner had tagged with a temporal term. We began looking for images tagged with 'christmas', and then five terms for seasons (spring, summer, fall, autumn and winter). We repeated this for a total of 56 temporally- or holiday-related search terms, including halloween, christmas, passover, eid, bastille, guy fawkes, pride, ski week, and so on. Each pattern was then visualized in Tableau in several different ways—by day of the year, by day of the week, and so on—to help identify different patterns. We treat as ground truth the date the photograph was taken as recorded in the image's EXIF metadata. This makes the assumption, of course, that this date is accurate. As shown in [9], 93% of geotagged images on Flickr had accurate timestamps within 48 hours. This tends to be less of a problem for mobile photos, as phones generally set their time from the network.

FINDINGS

In this section we will explore findings from both studies, which we will tie together when forming implications for design. Through analyzing data from search queries as well as from tags, we are able to develop a broader understanding of people's perceptions of temporal terms for personal images.

Interactive Survey

From the online survey, we gathered users' intended search queries for 1,492 images. For each image, we also asked if the participant had taken the photo themselves, and 1135 (76%) had been captured by the person taking the survey. The majority of photos were taken between 2004 and 2015 according to the EXIF metadata, with the median age photo taken in 2011. A histogram of dates captured can be seen in Figure 1. Most participants therefore had a large range of

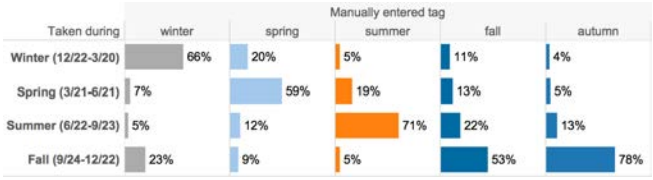

Figure 2. Photos tagged with particular season names do not necessarily occur in the canonical season. Note, especially, that 22% of Winter images actually were taken in the Fall, and 21% of images tagged Spring were taken in the Winter.

images from several years represented in the images selected for the survey.

Looking at the queries that participants provided, only 162 of the 1492 queries (10.8%), succeeded in finding the target image via the Flickr search API. We were particularly interested in examining queries with temporal terms in them, to see how participants remembered and expressed the dates of the various events. Twelve percent of all queries contained a temporal term: e.g. a year (69% of these queries), season (36%), holiday (18%), or month (11%). Note that some queries had combinations of these, such as 'December 2011'.

However, frequently these temporal terms did not match the EXIF date on the photo, and thus resulted in the search failing to find the target image. In looking at all queries with year-based terms, 27% of the time the year did not match the captured date on the target image. Month-based queries fared worse, with 33% of the months mentioned in the queries not matching the image data. Finally, season-based queries had the largest rate of error, with 40% of queries with season-based terms not matching the calendar definitions of the seasons.

Examining season-based queries illustrates how search systems can better adapt to people's memories of time. For example, all queries with the term 'winter' had EXIF dates between December 1 and March 31. However, using the calendar definition of winter (December 21st–March 20) as most systems do today, 20% of these images did not match the queries.

Data Footprints

Working with large-scale data confirmed our impressions from the small-scale study. For example, Figure 2 shows an analysis of when photos tagged with different seasons were actually taken. We are comparing manually tagged images (including the synonyms 'autumn' and 'fall' for completeness) with the four seasons as defined by the solstices and equinoxes for the Northern Hemisphere. Antipodean seasons may legitimately explain the 5% of images tagged 'winter' which were taken during the summer, or, similarly, the 5% of images tagged 'summer' taken during the winter. But we believe that the off-by-one cases, like fall photographs labeled winter, suggest the photo represents the *experience* of winter, regardless of the arbitrary and objective reality of the calendar.

One issue in using large datasets as guidance to inform translation of temporal terms to specific dates is the long tail of responses. For example, Figure 3, represents every photograph tagged with the word 'thanksgiving' on Flickr. The six days between November 22 and 27th inclusive, the darkest blue area,

Figure 3. The relative frequencies by day of the year of photographs on Flickr tagged 'thanksgiving'. Note the dark blue area showing American Thanksgiving, and the much lighter gray area in early October representing Canadian Thanksgiving.

covers 65% of the photos. Expanding that range to November 15–30 covers 83%. Expanding to all of November covers 85%, and including October (and thus Canadian Thanksgiving, in dark grey in early October) brings the total to 90%. But that means that 10% of all photos tagged thanksgiving are outside of this range. Every grey square in Figure 3 represents a total of a minimum of 40 photographs taken on that day between 2003 and 2014 inclusive, uploaded to Flickr and tagged 'thanksgiving' with the only white spaces being days that don't exist, like February 30th or April 31st. Repeated manual verification of public photos tagged thanksgiving on arbitrarily chosen dates confirms is not an error: it merely reflects the very long tail of such datasets. Such photographs tagged thanksgiving may include pumpkins or turkeys, autumnal leaves or cornucopias, all images culturally associated with the holiday. This underlines the need for careful search algorithm design to support such diverse practices.

The tagged data reveals that not all temporal search terms are quite so complicated. Some holidays are celebrated and captu photographically on a single day throughout the year, such as Canada Day, on July 1st, or Boxing Day, on December 26th. Treating searches for those terms as searches for photographs on the day itself, or one day on either side to allow for complications of time zones and inaccuracies, will generally return the majority of desired images. Others have more complicated temporal patterns. The lunar festivals of Easter, Passover, Chinese New Year, Eid Al-Adha, Eid Al-Fitr and so on move in a regular (if complicated) way when expressed on the solar calendar, as shown in Figure 5.

Even a term that one would expect to have minimal temporal change can turn out to have temporal patterns: photographs tagged 'birthday' are primarily taken on Saturdays (38%), then Sundays (24%), and then Fridays (12%). The remaining 26% is evenly distributed among the other days of the week as seen in Figure 5. This suggests, for example, that if a system was

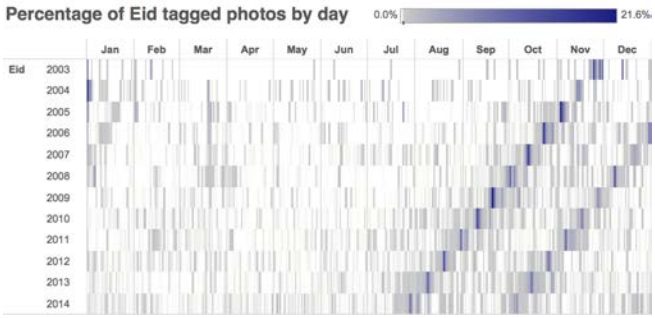

Figure 4. The Hijri (Muslim) Calendar is based off a lunar cycle with no leap correction. Here we see two "Eids" (holidays), Eid Al-Fitr and Eid Al-Adha, shifting dates as it moves across the Gregorian, solar calendar.

Figure 5. The relative frequencies by day of the week of photographs tagged 'birthday'. Birthday photos tend to be taken during weekends.

searching for a birthday photograph then there would be value in expanding the search term to nearby weekends.

Similarly, Halloween falls on October 31st. But Figure 6 shows photographs of the event are clustered on the Friday, Saturday and Sunday *before* Halloween. While 28% of the images tagged Halloween are taken on October 31st, the five days after Halloween contain only 10% of the images compared to 30% in the five days before Halloween.

Finally, to address the question posed in the title of the paper, when is Christmas? More photos tagged 'christmas' are taken on December 25th than any other day: 19%. Christmas Eve is a close second, at 12%. In other languages, this difference practically goes away: 9.2% of photos tagged 'noel' are taken on Christmas Eve, and 9.6% are on Christmas; 'navidad' photos are 11.3% on Christmas Eve and 12.0% on Christmas. But Christmas photos are taken throughout December. Looking at dates with at least 1% of the photos tagged 'Christmas' we find that every day from December 1st to January 1st hits that definition, with December 2nd barely scraping in; that makes 32 days of Christmas.

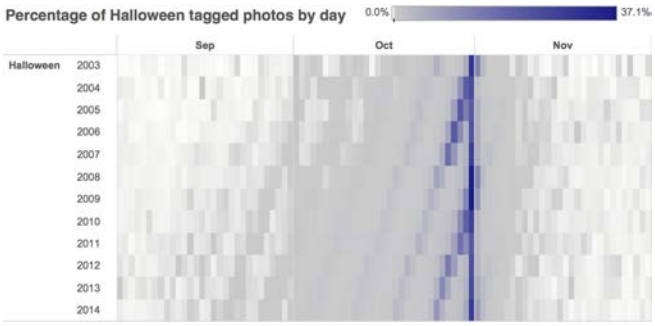

Figure 6. Photos tagged with Halloween are more likely to occur in the week before the holiday, with higher density on the weekend before.

IMPLICATIONS FOR DESIGN

This work highlights the importance of understanding people's perceptions of temporal aspects of their own events when designing organization and search systems. We have seen, through analyzing people's own search queries on their personal photos, and by examining the tags that people apply to their photos, that temporal terms often extend far beyond the calendar definitions in a type of vocabulary mismatch [3].

Beyond photos, other systems where people refer to dates in finding content (web search, calendar search, social network search, etc.) should consider people's understanding of time phrases and not stick to canonical dates. Specifically, search systems should extend ranges for season-based terms to extend more broadly beyond the strict calendar range. Especially in places with strong seasons, a snowy December 3rd definitely feels like winter, and our data shows that people do remember these events in this way. Currently, Flickr defines Christmas to be Dec 24–25, while Google+ defines it as Dec 23–25, and Facebook rewrites the query term 'Christmas' to 'December' and does not seem to rank based on proximity to the 25th.

Search systems should also show results for adjacent temporal ranges. For example, searches for 'July' should have easy ways to see the June or August results. Year-based searches such as '2011' should likely also include results from 2010 and 2009 for cases where people mis-remembered and need to easily expand their query to find the target result.

Finally, returning the title of this paper, holidays should not be seen as only specific dates. Some holidays, especially ones that take place over longer time periods or involve travel such as Christmas, can often include events over an entire month, and our search systems should support this broader context of the term beyond just December 24th and 25th. Likewise, one might approximate Thanksgiving as the entire 4-day weekend or even the entire week around the event, using weighted algorithms to represent differing probabilities.

CONCLUSION

We have analyzed the ways that participants naturally thought of temporal terms when searching for or tagging photos in their own collections. Through showing that memory of years, months, and seasons can be off of standard definitions up to 40% of the time, we have demonstrated the need for alternative solutions to help people find content when using temporal information. We have provided several clear implications for design for those who are building content organization and search applications.

Future work can explore these same types of errors and misremembrances on other types of metadata including locations, people in the photos, or social history of the photo (e.g. who it was sent to). Additional work can explore these same temporal memory issues for other types of content including searching for email, files, or news. As search becomes more important across content types in our lives, and as the amount of content that we have to search increases, understanding human perception and memory of temporal terms will only become more important.

REFERENCES

1. Morgan Ames, Dean Eckles, Mor Naaman, Mirjana Spasojevic, and Nancy Van House. 2010. Requirements for mobile photoware. *Personal and Ubiquitous Computing* 14, 2 (2010), 95–109.
2. Frank Bentley and Crysta Metcalf. 2006. Flexible Views: Annotating and finding context-tagged mobile content. In *Ubicomp 2006 workshop on Pervasive Image Capture and Sharing*. Citeseer.
3. W Bruce Croft, Donald Metzler, and Trevor Strohman. 2010. *Search engines: Information retrieval in practice*. Addison-Wesley Reading.
4. Marc Davis, Simon King, Nathan Good, and Risto Sarvas. 2004. From context to content: leveraging context to infer media metadata. In *Proceedings of the 12th annual ACM international conference on Multimedia*. ACM, 188–195.
5. David Frohlich, Allan Kuchinsky, Celine Pering, Abbe Don, and Steven Ariss. 2002. Requirements for photoware. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*. ACM, 166–175.
6. Mary Meeker. 2014. 2014 Internet Trends. (2014). http://www.kpcb.com/internet-trends
7. Mor Naaman, Susumu Harada, QianYing Wang, Hector Garcia-Molina, and Andreas Paepcke. 2004. Context data in geo-referenced digital photo collections. In *Proceedings of the 12th annual ACM international conference on Multimedia*. ACM, 196–203.
8. Meredith Ringel, Edward Cutrell, Susan Dumais, and Eric Horvitz. 2003. Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proc. Interact*, Vol. 2003. 184–191.
9. Bart Thomee, Jose G Moreno, and David A Shamma. 2014. Who's Time Is It Anyway?: Investigating the Accuracy of Camera Timestamps. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 909–912.
10. Steve Whittaker, Ofer Bergman, and Paul Clough. 2010. Easy on that trigger dad: a study of long term family photo retrieval. *Personal and Ubiquitous Computing* 14, 1 (2010), 31–43.