Image Annotation Using Personal Calendars as Context

Andrew C. Gallagher
Research Laboratories
Eastman Kodak Company
Rochester, NY
andrew.gallagher@kodak.com

Carman G. Neustaedter
Research Laboratories
Eastman Kodak Company
Rochester, NY
carman.neustaedter@kodak.com

Liangliang Cao
Beckman Institute
University of Illinois
Urbana, IL
cao4@illinois.edu

Jiebo Luo
Research Laboratories
Eastman Kodak Company
Rochester, NY
jiebo.luo@kodak.com

Tsuhan Chen
Dept. of Electrical and
Computer Engineering
Carnegie Mellon University
Pittsburgh, PA
tsuhan@cmu.edu

ABSTRACT

In this paper, we introduce the idea of using the context of a personal calendar for labeling photo collections. Calendar event annotations are matched to images based on image capture time, and a Naïve Bayes model considers features from the calendar events as well as from computer vision-based image analysis to determine if the image actually matches the calendar event. This approach has the benefit that it requires no extra annotation from the consumer, since most people already keep calendars. In our test collections, 36% of personal images could be tagged with a label from a personal calendar. Note that our preliminary results represent a lower bound on the performance that is possible because all of the system components are expected to improve over time. As people migrate toward digital calendars, we can also expect more consistency in their calendar labels, which should improve the annotation accuracy.

General Terms

Algorithms, Human Factors, Experimentation

Keywords

Calendar, image annotation, content-based image retrieval

1. INTRODUCTION

With the advent of digital photography, consumers today are capturing images at a rate never before achieved in history. As consumers' image collections grow, the practical problem of finding images becomes apparent. Computer vision researchers are actively designing algorithms for annotating images with meaningful annotations. When successful, these annotations will allow consumers to search and browse their image collections in a more efficient manner and in new ways. We are interested in improving consumers' access to their image collections. We assume that the queries a consumer would make are related to the attributes that the consumer remembers about the images. Naaman [5] found that the most remembered attributes of an image are indoor/outdoor, identities of people in the image, location, and event.

With this in mind, the task of annotating images with relevant information is still difficult. Some applications exist (e.g., www.flickr.com, Adobe Album) that allow a user to manually tag images with relevant labels. Of course, these manual solutions are labor intensive and a more automatic solution is sought. Unfortunately, the current state of computer vision algorithms is not sophisticated enough to provide image annotations that are ideal for search and

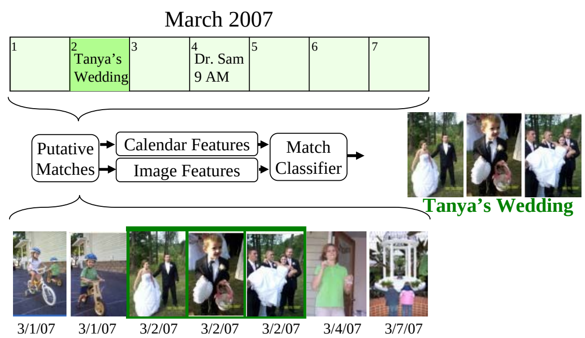


Figure 1: Personal calendar information is useful for annotating images. Based on coincidence between the calendar event date and the image capture times, putative image-event matches are found. For example, the calendar event "Tanya's Wedding" matches with the third-fifth images from the collection. Calendar and image features are extracted and the match classifier determines whether the images match the event, successfully tagging those relevant as "Tanya's Wedding." Likewise, the putative match between "Dr. Sam 9 AM" and the sixth image is a non-match that is not annotated. Meaningful tags are generated from calendars with no extra effort!

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

retrieval. Researchers have shown that specific objects (such as faces, pedestrians, and cars [2, 3]) can effectively be located, but these objects are not necessarily related to queries that a consumer might make. For example, a consumer is generally more interested in finding "images of our vacations" or "images from our wedding" than "images containing cars."

Many researchers are addressing the special issues associated with recognizing people in image collections. Furthermore, when the locations of images are known (e.g. through GPS tagging), the event location is also known. Unfortunately, this technology is not widely available and most consumers do not have geotagged images.

We propose a method for using personal calendars for labeling consumer images. To provide semantically meaningful image labels, we must employ all possible context related to the capture of the images. We observe that calendar entries (e.g., "Meet Jenny for Picnic" on May 12, 2008) provide a great deal of context for labeling images captured at a coincident time.

Our novel approach has the following advantages:

1. Calendar entries contain high-level semantic information such as Event, Location, and Names of people, which as Naaman showed are useful for consumer or personal image retrieval [5].

2. Consumers already keep calendars. Although the family calendar is generally kept on paper, it is likely it will eventually migrate to digital form [6], where the information will be easily available for our photo-labeling application. We glean information for labeling images without requiring extra effort from the consumer.

2. A CALENDAR AND IMAGE MODEL

We would like to build a model that represents the relationship between consumers' calendars and the images that they capture. This model allows us to determine when the calendar event is an appropriate annotation for one or more images from the collection. By using a consumerŠs calendar labels as potential tags, we are able to tag a consumerŠs images with their own words, likely increasing the chance that these tags will serve as useful query terms.

Figure 1 illustrates our approach. We assume access to both a consumer's calendar (top left) and image collection (bottom right). Putative matches are found between calendar events and images from the collections based on coincidence between the event and image capture times.

For many of these putative matches, the calendar event is a relevant tag for images captured at the corresponding time. For example, a wedding or birthday party is often noted on a calendar and corresponds to images in a collection. However, not all putative matches represent appropriate image tags. To understand the reasons why, we must understand how calendars are used in a culture [6]. Not all calendar notations represent events in which the person will participate. For example, a calendar notation of "Mom's Birthday" might simply represent a reminder to call Mom (rather than attend a party), and photos captured on that day might have no connection at all with "Mom's Birthday."

In summary, a calendar is a complex mixture of image-capturing events and non-image capturing events. When a putative match is found between a calendar and an event, our match classifier is used to classify whether the images and calendar events are a match, based on image and calendar event features. Upon determining that a match exists, the calendar event is used to tag the images. For example, in Figure 1, a correct result from the match classifier allows us to correctly tag the three wedding images as "Tanya's Wedding" and prevents us from tagging the sixth image as "Dr. Sam".

Type	Name	Dimension
Event	`Multiday`	1
	`Preprinted`	1
	`TimeAssociated`	1
	`Words`	745
Image	`EventClass`	11

Table 1: The features used by out match classifier to determine if a putative match between an image and a calendar match is, in fact, an actual match.

3. FEATURES AND CLASSIFICATION

A putative match between a calendar event e and a set of images $\mathbf{I}$ is denoted $P_{e\mathbf{I}}$. Putative matches are found with the following simple method: Each calendar event e has an associated start and stop time. When an event has no associated start and stop time, it is assumed that the event duration is for the entire 24-hour day. If an event is noted with a single clock time, it is assumed this time marks the start of the event. Note that the event can have a duration of several days (e.g., a vacation). Meanwhile, images are clustered employing their image capture times using a method similar in performance to [4]. When the image capture time of an image cluster $\mathbf{I}$ falls within the duration of an event e, a putative match occurs. Features $\mathbf{F}_e$ are extracted from the calendar event e and the images $\mathbf{I}$ from the putative match $P_{e\mathbf{I}}$. Table 1 provides a summary of the features.

3.1 Event Features

In addition to the words that comprise a calendar event e, we include three Boolean features. `Multiday` indicates whether an event duration spans multiple days (true) or not (false). `Preprinted` indicates whether an event was printed onto the calendar (e.g. a holiday) or added manually by the consumer. `TimeAssociated` is true when the consumer indicates a precise start or stop time with the event.

The words associated with an event also provide an indication of whether images will be associated with the event. Even without observing images from a putative match, we can guess that "Church Picnic" might represent an image-capturing event, but "Oil Change" probably does not. When parsing the event words, we also determine if the word represents a first name, using data from the U.S. Social Security Baby Name Database [7] and consider all first names as equivalent. After collapsing first names, we have 745 unique words among the six consumer calendars (Section 4).

3.2 Image Features

We perform image analysis on the images from a cluster I, using the method of [1]. Rather than directly learning the relationship between every calendar event word and image appearance (which essentially require the solution of computer vision), we chose to learn the relationship between the event categories of [1] and the likelihood that a putative match is in fact an actual match.

In [1], an event label is found for all of the images in an image cluster ("subevent"), after the timestamp information is used to form the clusters. To briefly summarize, the event label is found for a cluster of images by using conditional random field (CRF) models to exploit the correlation between photos based on a hierarchy of (1) time-location constraints and (2) the relationship between collection-level annotation (i.e., events) and image-level annotation (i.e., scenes). From each image, low-level vision features are used by a suite of trained SVM classifiers to generate the initial event labels that are revised by enforcing the above-mentioned constraints. The labels are constrained to the following 11 events (excluding the Null event):

`BeachFun`	`BallGames`	`Skiing`	`Graduation`
`Wedding`	`Birthday`	`Christmas`	`UrbanTour`
`YardPark`	`Family`	`Dining`	

While it is indeed unlikely that a consumer will annotate their calendar with terms like "BeachFun", we are not necessarily trying to use these categories to label images. Rather, we propose learning the relationship between these categories and $P(M_{e\mathbf{I}})$, the probability that putative match $P_{e\mathbf{I}}$ is an actual match.

3.3 Classification

The event and image features for a putative match $P_{e\mathbf{I}}$ are concatenated into a single feature vector $F_{e\mathbf{I}}$. A Naïve Bayes model is used to represent the joint distribution between the features $F_{e\mathbf{I}}$ and the class variable $M_{e\mathbf{I}}$. We experimented with using different combinations of the putative match features. In our experiments (Section 6), we demonstrate the validity of our approach using the personal calendars and image collections from six consumers.

4. USER STUDY METHODOLOGY

We first conducted a user study to first test whether or not calendar event labels could be used as a reliable tool for automatically tagging images. We recruited six participants (five female, one male) using a study recruitment agency. Participants had a variety of occupations and family compositions (e.g., families with young children, parents with adult children). All participants were representative of typical digital photography consumers and provided their 2007 family calendar and images taken during the same calendar year. Prior to each session, the participant's calendar was manually digitized to create a data file containing a list of calendar events and associated information.

Each study session included two stages:

1. **Image and Event Matching:** Participants were shown each of their calendar entries along with any images taken on the same day as the event (e.g. the image set of the putative match $P_{e\mathbf{I}}$.) Participants then selected photos from the set that were taken at the event listed. This selection process repeated for each putative match.

2. **Quality Assessment:** Participants were shown each image that they had previously associated with an event and asked to assess how well the calendar entry (e.g., the text written on their calendar for that event) described the image. This was rated on a five-point Likert scale [8] from Not Very Well to Very Well. To avoid user fatigue, if participants had more than 50 matching images, then 50 randomly selected images and matched calendar annotations from their set were shown for evaluation. Four participants rated 50 images, one rated 16, and another rated 17. Participants could optionally type in a better label for each image.

	P1	P2	P3	P4	P5	P6	Total
Images	1432	201	10	569	331	99	2642
Events	310	199	433	66	134	106	1248
Putative Matches	1728	261	16	261	406	156	2831
Actual Matches	1088	180	16	80	174	18	1556

Table 2: Image and Calendar Collection Statistics.

	P1	P2	P3	P4	P5	P6
N	50	50	16	50	50	18
Mean	3.2	2.4	2.2	3.0	4.8	3.6
Median	4	2	3	3	5	4

Table 3: Annotation Quality Assessment.

4.1 Image and Event Matching

Table 2 shows a summary of the data collected from our six participants. A total of 2642 images and 1248 events were collected. Out of 2831 putative matches between calendar events and images, 55% (1556) were actual matches. Further, we note that in the ground truth, 960 out of the 2642 images (36%) could be labeled with at least one calendar event.

4.2 Label Quality

Participants rated the quality of the calendar labels that they associated with each image in the second stage of the user study. Table 3 shows the number of images rated by each participant (Row 1), the average score given across all the images they assessed and standard deviation (Row 2), and the median score (Row 3). Ratings were on a five-point Likert scale from 1 (the label does not match the image very well) to 5 (the label matches the image very well). On average, most participants rated the quality of the calendar labels as being satisfactory for describing the images: five of six participants' average and median ratings were between 2 and 5.This shows that if this technique is to be readily used by people, its quality will depend heavily on what people write on their calendars. However, it is quite remarkable that many calendar events were found to be good image annotations even though the calendar annotations were never made with that application in mind. It is even more encouraging to note that over time, people may adapt what they write on their calendar if they knew that the labels could be used to tag images.

5. CLASSIFYING PUTATIVE MATCHES

We train our Naïve Bayes model to predict $P(M_{e\mathbf{I}}|F_{e\mathbf{I}})$, the probability that putative match $P_{e\mathbf{I}}$ is an actual match given the features. The model is trained by leaving out one putative event-image match and training on the remaining matches from the same participant. We also explored leaving out one participant's collection and training on the other five collections. While this method still surpasses the prior, the results are not as good (area under the curve is 0.66). We believe this is because with only six participants, it is not easy to learn a general calendar vocabulary.

Combining Image and Event Features: Our results are summarized by performance-recall curves in Figure 2. The areas under the performance curves are included in the legend. The noisy nature of the curves is a result of using

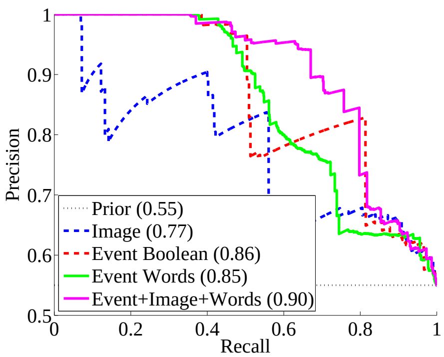


Figure 2: Classifying a putative calendar event-image match as an actual match or a non-match can be accomplished with image features, event features, or both. When image and event features are combined, performance is best.

variables with a relatively small number of discrete states, and because groups of images (within a cluster) have similar features. Compared with the prior match probability of 55%, performance is significantly improved by using the Boolean event features, event words, or the image features. As expected, the best performance is achieved by merging both event and image features. Figure 3 shows example images from one of our participants. The top row shows images from a wedding (left) and New Year's Day (right) correctly matched with the associated calendar entries. The middle row shows two images correctly not matched to a birthday party (left) and a dentist appointment (right). The bottom row shows collectibles (left) and an outdoor scene (right) incorrectly matched to a hockey game and dentist appointment, respectively.

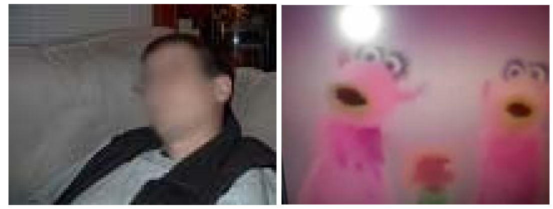
Event Words: Many calendar words are common and contain luttle information about whether a putative match is an actual match. However, the presence of any of a small number of (infrequent) key words W are features that can predict $P(M_{e\mathbf{I}}|W)$ with high confidence. Among the calendar words with the highest $P(M_{e\mathbf{I}}|W)$ are "Party", "Shower", and "Camping", while words with low values for $P(M_{e\mathbf{I}}|W)$ include "Dr.", "Dentist", "School", and "Scouts". This aligns with our intuition that certain events are more likely to be photographed than others.

In an additional experiment, we compared the event label words with the image-based event classification from [1] for putative matches. When they share a common word (e.g. in our test set, matches occurred for "Wedding", "Graduation", and "Birthday") it is extremely likely that the putative match is an actual match ($63/67 = 94\%$). This highly accurate feature improves over our results in Figure 2. In addition, the 94% accuracy represents a significant improvement in accuracy over vision-only classification in [1]. However, it is rare that they match (only 67 occurrences among the 2831 putative matches in our data).

6. CONCLUSIONS

In this paper, we have introduced the idea of using the context of a personal calendar for labeling the corresponding photo collections. Calendar annotations are matched to images based on image capture time, and a Naïve Bayes model considers features from the calendar events as well as

Images **correctly matched** with calendar events.



Images **correctly not matched** with calendar events.

Example images **incorrectly matched** with calendar events.

Figure 3: Example images illustrating our algorithm results. The Naïve Bayes model is used to detect when calendar labels are appropriate annotations for putatively matching images.

from computer vision-based image analysis.

This approach has the benefit that it requires no extra annotation from the consumer, since most people already keep calendars. It would require that people use digital calendars, as opposed to paper ones. This is a likely future of family calendaring [6]. Furthermore, our results represent a lower bound on the performance that is possible. As people migrate toward digital calendars we can expect more consistency in their calendar labels and this in turn should improve accuracy of photo annotation.

Our ongoing work involves expanding our test set to allow us to better learn the interactions between the images and calendar event features.

7. REFERENCES

[1] L. Cao, J. Luo, H. Kautz, and T. Huang. Annotating collections of photos using hierarchical event and scene models. In *Proc. CVPR*, 2008.

[2] N. Dalal and B. Triggs. Histograms of orientated gradients for human detection. In *Proc. CVPR*, 2005.

[3] M. Jones and P. Viola. Fast multiview face detector. In *Proc. CVPR*, 2003.

[4] A. Loui and A. Savakis. Automated event clustering and quality screening of consumer pictures for digital albuming. *IEEE Transactions on Multimedia*, 2003.

[5] M. Naaman. *Leveraging Geo-References Digital Photographs*. PhD thesis, Stanford University, 2005.

[6] C. Neustaedter. *Domestic Awareness and the Role of Family Calendars*. PhD thesis, University of Calgary, 2007.

[7] U.S. Social Security Administration. Baby name database. http://www.socialsecurity.gov/OACT/babynames.

[8] Wikipedia. Likert scale. http://en.wikipedia.org/wiki/Likert_scale.