

Extractable Mobile Photo Tags

Ramesh Jain
Department of
Computer Science
University of
California, Irvine
Irvine, CA 92697
jain@ics.uci.edu

Mingyan Gao
Department of
Computer Science
University of
California, Irvine
Irvine, CA 92697
gaom@ics.uci.edu

Setareh Rafatirad
Department of
Computer Science
University of
California, Irvine
Irvine, CA 92697
srafatir@ics.uci.edu

Pinaki Sinha
Department of
Computer Science
University of
California, Irvine
Irvine, CA 92697
psinha@ics.uci.edu

ABSTRACT

Mobile phones are resulting in a major shift in how people shoot photos. Just a little more than a decade ago consumer behavior was *plan-shoot-process-share-organize-reflect*; but rapid proliferation of mobile phone cameras have resulted in *shoot-share-forget* behavior. This trend will be replaced soon because photos are more important than that – people treasure their memories in visual form. Fortunately, a plethora of sensors combined with access to powerful Web may allow effortless *organize and reflect* environment without much, if any, cognitive load on the consumer. We propose new approaches for determining attributes that we call *Extractable Mobile Photo Tags* (EMPT) for processing and organizing photos and videos on mobile phones. We present approaches to populate EMPT and use it for applications.

Author Keywords

Personal Photo Organization / Management, Contextual Information, Extractable Mobile Photo Tags.

ACM Classification Keywords

I.2.4 [ARTIFICIAL INTELIGENCE]: Knowledge Representation Formalisms and Methods – *Semantic Networks*;

H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems – *Evaluation/methodology*.

General Terms

Algorithms, Management, Experimentation, Human Factors.

INTRODUCTION

Photo management as recently as 10 year ago was a very different problem than it is today. Our techniques are still from the old world, however. At one time cameras were used to take photographs that represented intensity values at a point on the film, or at a pixel on a CCD array. Emergence of digital cameras, particularly those in smart phones, has radically changed the nature of photography

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MLBS'11, September 18, 2011, Beijing, China.

Copyright 2011 ACM 978-1-4503-0928-8/11/09...\$10.00.

and photo-taking habits of people. The radical changes have come along two dimensions. First, due to the ease or capturing photos and relatively no cost associated with photos, people take many photos and store them. Second, these devices capture a host of contextual information, commonly called metadata, like time, location, camera parameters, and voice tags along with the media itself. There are several sources that feed different kind of contextual information to a phone ranging from location information to calendar, contacts, and information in clouds, as shown in Figure 1. When a photo is taken from this camera the system can effortlessly add advanced meta information that we will call *Extractable Mobile Photo Tags* or **EMPT** to the photo header.

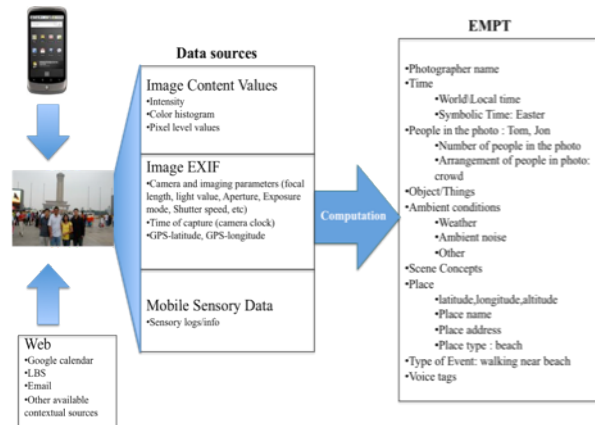


Figure 1: Data sources from phone and Web are used to populate EMPT fields.

In fact, a photo is no longer just an array of intensity values; it is experiential data associated with an event during which the photo is captured. Cameras capture significant metadata associated with photos and much more can be inferred from other sources. This metadata is much less noisy compared to the human induced tags in online image sharing platforms. However, the challenge lies in correctly interpreting this multi-modal data to determine useful attributes for organizing and annotating photos for easy retrieval, reliving, and reflection. In this paper we discuss multiple approaches being developed in our lab to populate EMPT for a photo. We also present our efforts to develop summarization approach for showing summary of photos

on the phones so that the visualization and management of a large number of photos could become an enjoyable task.

EXTRACTABLE MOBILE PHOTO TAGS

For the photos captured by smartphones like the Android, it is possible to add several semantic fields to each photo. These semantic fields could be added when the photo is captured using background processing for each photo. We started identifying such fields, listed below, based on the availability as well as efficacy in photo management and search.

Photographer Name: This information will be obtained from the ownership of the camera.

People in Photos: Several fields will be added here: Numbers and Names of People, Portrait photos or crowd.

Place: In addition to lat-long-alt; Type of place; Name of place.

Event: Using calendar and event detection techniques, the system will detect and store Type and Name and other details of event.

Environment: Based on location and time, system will determine and store environment conditions such as Weather class (Cloudy, rain, ..), ambient noise (loud, ...) , etc.

Objects/Things: Using computer vision techniques combined with strong context, determine objects such as Pets and other favorite things in photos.

Scene Concepts: Modify current computer vision concepts of scenes [5] (beach, outdoor, city, mountains, ..) and develop context guided techniques for learning these concepts in photos and use them as enumerated concepts.

Time: In addition to the clock and data time, one should also consider storing personal (birthdays, anniversaries, ..), local festivals (Chinese New Year, Deepavali, Easter, ..), and other significant symbolic time indicators.

For the fields that the value can not be inferred reliably, 'UNKNOWN' will be stored so that a user may supply that value if desired.

In the following, we will briefly discuss research projects to populate EMPT and use it for various applications.

RECOGNITION OF EVENTS AND SIBEVENT STRUCTURE

Given photos with EXIF metadata for an event, we partition them into its sub-events. We use domain event ontology corresponding to the type of the event, instantiate the domain ontology using information available for the event (i.e. time, location, participating people), and augment the ontology instance with all available information related to

the context of the subevents. Domain Ontologies are nothing but formal conceptual models at the “semantic” level that are independent from lower level data models [6]. High-level semantics (e.g., an event that a photo covers) are linguistic descriptions and a linguistic description is almost always contextual [7].By augmentation, we mean associating values to an individual-event context. This instantiated contextual model, called R-Ontology, is shown in Figure 2. R-Ontology is then used to partition the given photos.

The corresponding domain event ontology describes the underlying event and its subevents in terms of their parthood relationship (i.e. *subevent-of*), relative temporal relationships (i.e. *previous-event*, *next-event*, *started-by-event*, *finished-by-event*), environment, scene concepts, and object/things.

R-Ontology is derived from such event model. The relationships in R-Ontology are described with properties such as negation, transitivity, etc; subevent-of, next-event, previous-event are examples of relationships with transitive property. In addition to that, we consider some inference rules, for instance, consider the following case described in the domain ontology:

previous-event (e1,e3), next-event (e1,e2) → started-by(e1,e3.end), finished-by(e1,e2.start).

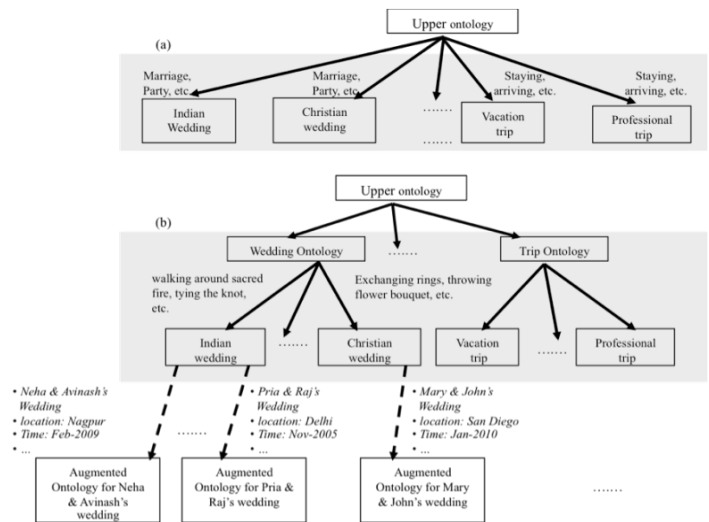


Figure 2: R-Ontology

In the first line *e1* is a event that is temporally and relatively bounded by time-intervals of events *e2* and *e3*. Relationship *next-event/previous-event* indicate the very next/previous event (temporally) after/before the occurrence of *e1/e2* respectively. In the second line, predicates are inferred from the first line that shows the predicates in the corresponding domain ontology. This shows that our framework supports temporal proximity by translating relative temporal relations into absolute relations in R-Ontology.

Partitioning is conducted by function f as follows:

$$\begin{aligned}
 f : (P, e, O) &\rightarrow H, & (1) \\
 s.t. H &= \hat{E}(\langle Pi, sej \rangle), & (1.1) \\
 Pi &\hat{I} P, sej \hat{I} R\text{-Ontology}, \\
 &" Pi \text{ associate}(Pi, \{sej\}); \\
 f &: f_{extract} * f_{cluster} * f_{match} & (2)
 \end{aligned}$$

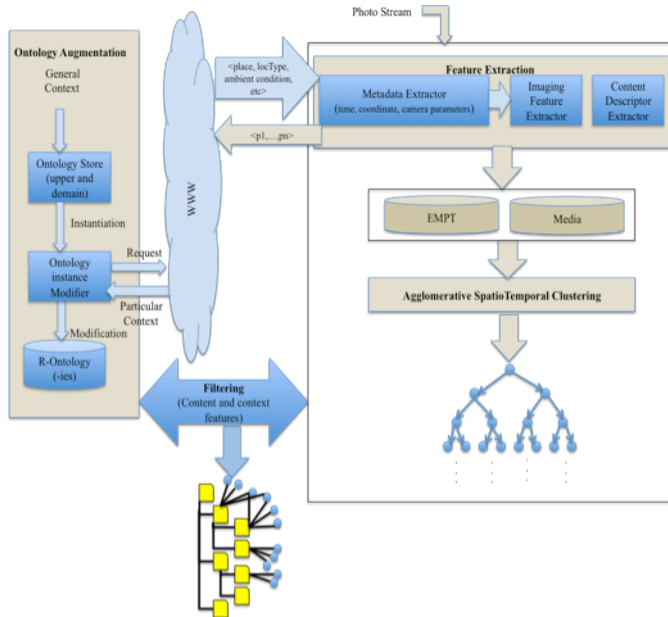


Figure 3: General Architecture

In (1), P is a set of photos given for event e , and O is a lookup table with two fields *name*, and *path* for available event domain ontologies in our framework. The lookup table is searched by the name of ontology (that matches the type of e) to get its URI. Function f that is a complex function (2) associates each P_i (a subset of P) with a set of subevents ($\{sej\}$); if this set is empty at the end of f , the corresponding field in the image is set to UNKOWN that means no subevent in the underlying R-Ontology covers subset P_i . The final result is a set of pairs (H) indicated in (1.1).

Each P_i is initially represented as a tuple with the following fields: $\langle id, interval, bounding\text{-}box, children\text{-}Ids \rangle$ that means it belongs to a hierarchical structure. The hierarchical structure is generated from running an agglomerative spatiotemporal clustering ($f_{cluster}$) on P --using time and location from EXIF metadata-- such that none of the direct children of P_i overlap the other; then each cluster is matched to some subevent/(s) according to its/their absolute time and location upon availability.

Finally, f_{match} selects a subset of images from those in the matched clusters to a subevent, by applying a set of constraints. These constraints are based on properties

environment, scene concepts, and object/things corresponding to the subevent. For this part, we obtained the corresponding properties of pictures used in constraint matching by $f_{extract}$, before f_{match} begins.

Figure3 shows the general architecture for this framework. Figure4 shows some results for a vacation trip. Given all the data that is available in mobile phones, we believe that this approach can be very effective.

For mobile smart-phones, we have initiated extending the idea of using R-Ontology to **real-time** scenario. The information detected by R-Ontology includes the actual event that covers a particular photo; the actual event may belong to an arbitrary level in subevent structure that is associated to the high-level event. All other context information comes with the actual event.

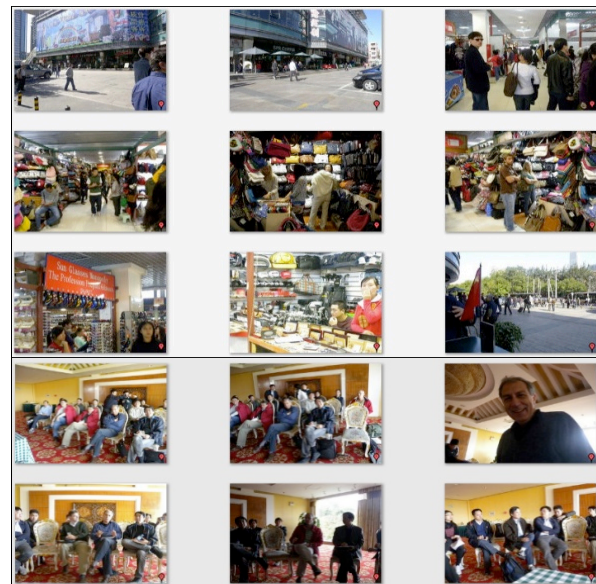


Figure 4. (a)Top 3 rows:Shopping;
(b)Bottom two rows:Talk

WONDER WHAT

Motivation

Life events, such as exhibitions, music lives, festivals, tours, nightlife, sports and various community activities constitute important parts of our everyday life. Correctly recognizing such events allows us to draw proper inferences about people and objects related to the events. Event recognition by mobile visual search has a great number of applications. For example, it is useful for travelers to get immediate information about local events. To get quick knowledge about these events, by using a mobile visual search tool, travelers need only focus their cameras on events to query and receive relevant event information. Events detected based on primary spatio-temporal context and content in turn provide a secondary context for recognizing objects and people involved in the events. For instance, at a conference, to get the information of the

current speaker, you can simply take a picture of his presentation. Now given the photo, the visual search is applied to determine the event, based on which names and faces of relevant people can be detected, compared, ranked and returned to you. Besides, with the increasing amount of online social information, such as friend relationship and social event announcements on Facebook, this technique can be employed to recognize social events and friends as well.

We discuss a system that provides users with information of the public events that they are attending by analyzing in real time their photos taken at the event. Whenever a user wants to know about an event she is currently at, she only needs to take a picture of it. By examining the photo content together with the spatial and temporal data carried with it, our system automatically returns a ranked list of events with which the photo may be associated. Our approach has the following advantages. 1) Use of the system is very intuitive and requires no special efforts; 2) The system keeps a dedicated event database and index, and automatically constructs queries for users, which enables the delivery of exact event information; 3) Our system not only detects planned events, but also tries to discover concurrent events by analyzing real-time micro-blogs; 4) Different types of events do reveal distinct visual characteristics, so visual content is also taken into account to improve search results. As far as we know, there is no previous work that has addressed a similar problem.

Problem

We formulate the problem which serves as the basis for the following discussion.

Contextual Photo

A contextual photo is represented as a triple $p = (img, time, location)$, where img is the image content of the photo p , and $time$ is the timestamp when the photo p was taken, and $location = (latitude, longitude)$ corresponds to the geo-coordinate of the photo shooting location. In our problem, time and location jointly identify a unique spatio-temporal context under which the photo was created. A contextual photo p is an input to our system.

Event

We follow the proposal in [1], and denote an event as a tuple $e = (time, location, title, description, type, media)$. $time = (start, end)$ is the time interval during which e occurs. $location = (lat1, long1, lat2, long2)$ represents the geo-coordinates of the southwest and the northeast corner of the place where e takes place. Name of event e is stored as string in $title$, and the textual explanation of e is saved in $description$. Event type indicating the class and genre of events, such as performances, exhibitions, sports and politics, are stored in $type$. Media data associated with some events, such as posters, photos and videos, is kept in $media$.

Problem Formulation

Given a contextual photo p , an event ranking function H is represented as:

$H : p \times E \rightarrow R$ where set $E = \{e_1, \dots, e_n\}$ is the event space, and each $e_i \in E$ is an event as defined in 3.1.2. R is the event ranking value space. The value of $H(p, e_i)$ represents the likelihood that e_i is the event at which the photo p was taken. Now given an input contextual photo p and an event ranking function H , the event recognition problem is to return an ordered list of events (e_{i1}, \dots, e_{in}) , in which $H(p, e_{i1}) \dots \geq H(p, e_{in})$. In this work, we consider spatial, temporal and visual features in event ranking function H . Each feature is again a ranking function $h : p \times E \rightarrow R$. The final output ranking value is computed as $H(p, e_i) = t=13wht (p, e_i)$, where wt is the weight associated with feature ht . We will explain the details of these features and their combination in the following discussions.

Implementation

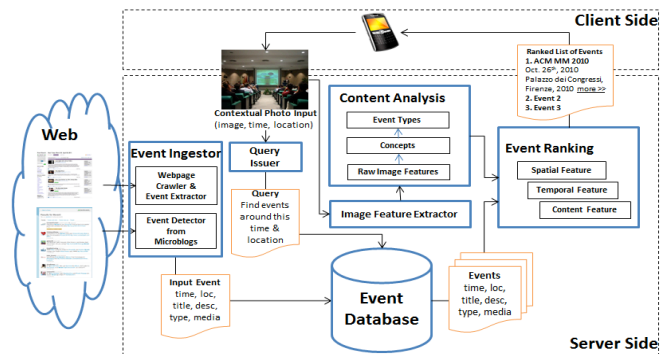


Figure 5. Architecture of WonderWhat system.

We present the system architecture in Figure5. Our system consists of the following major steps.

1. We create an event database, and ingest both planned and emergent events into it. Planned events, which are usually pre-declared online, are extracted from web pages or downloaded via web services that perform event integration. Emergent events, the occurrence of which is impromptu, are detected from Twitter.
2. After a user takes a photo for an event, her device creates a contextual photo containing the image content, time and location information. The device then sends this contextual photo to our system as a query.
3. First, given the time and location information in the contextual photo, a query is issued to the event database, which returns a list of related events.
4. Then the content analysis component analyzes the image content and returns the event type of the event captured in the photo. In this work, we model the relationship between event types and the raw visual features through a middle layer of visual concepts. We employed a learning based approach to perform the analysis, which consists of four major steps:
 - 1) Train concept detector;

- 2) Detect concepts from photos associated with different event types;
- 3) Train event type detector;
- 4) For each incoming photo, based on the models, decide which type of event the photo is most likely to be.
5. Both the event list from event database and the detected event type are given to the ranking component. The component considers spatial, temporal, and visual distances in the final ranking process.
6. Finally, a ranked list of events and their associated information are returned to the user and presented on her device.

Experiments

We conducted experiments on both Flickr dataset and a real event photo set shot in New York City.

Flickr Dataset

In this experiment, we verify the hypothesis that people do take photos at events. And by making use of the taking time and location of photos, we are able to match them to the corresponding events. We built the event database for events in NYC from year 2008 to March 2011. Also, we called the Flickr API and downloaded all the photos shot from year 2008 till March 2011. We matched the photos to the events in the event database. Figure 6 shows examples of matched events and photos. The left column details the events and the URLs where these events were extracted, and the right column lists the photos taken at the events.

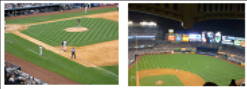




Collected Events	Flickr Photos
Title: New York Yankees vs. Boston Red Sox Time: 2009-05-05 19:05:00 Location: Yankee Stadium Type: Sports http://new.york.events/guide.com/qsus/90500.htm	
Title: 8th Annual Arab-American Street Festival Time: 2010-07-10 09:00:00 – 19:00:00 Location: Great Jones Street between Broadway and Lafayette Street, Manhattan Type: Festivals http://www.nyc.gov/portals/site/nycgov/menuitem.b0179b01de170747472ee1877089a0/index.jsp?&spic=contacts&EPR_C&epi=process&generic_proc=dlc_jsp&bean=0a1622145301&view=0ca&mode=process_view&page=0eventrowwise&range=day&selectedDate=7/10/2010&pg=0	
Title: A Rocket To the Moon Time: 2011-03-16 18:30:00 Location: Highline Ballroom Type: Concerts, Pop/Rock http://www.nyc.com/events/?ac=308&rome=03%2f16%2f2011&te=03%2f16%2f2011&sort=8&dir=asc&int=1&pg=1&list	
Title: Flower Show!! (@ Macysw/ 18 others) http://4sq.com/hY3oLt Time: 2011-03-26 20:48:41 Location: (40.74978786, -73.98842358)	
Title: Three centuries of red and white quilts in one (giant) room @ Park Avenue Armory Time: 2011-03-28 17:09:24 Location: (40.767616 -73.9662391)	

Figure 6. Examples of matched events and photos.

Real Photo Set

In this section, we test on real photo sets collected from 4 volunteers living in NYC. We asked each volunteer to hang around on streets in NYC during their spare time, in August and September 2010, and try attending some events that they discovered. They were advised to take as many pictures as possible at the event, and there were no

requirements on the subjects of these photos. The ranking result is depicted in Figure 7. The photo column shows a sample of pictures from each photo set, and the result column lists the top 5 ranking result for most pictures in the photo set. The last column provides the ground truth of these events. For event 1, 3, and 4, we correctly returned the information of the corresponding events on the first place in ranked lists. But for event 2, since the exact event was not stored in our database, our system returned a musical event in the Mitzi Newhouse Theater of Lincoln Center, which was a very close match.

Photos	Ranking Result	Ground Truth
	<ol style="list-style-type: none"> 1. Free Music Fridays 2. Target Free Fridays 3. Film Collection at MoMA 4. Drawings Collection at MoMA 5. Contemporary Art from the Collection 	Title: Free Music Fridays Time: 09/10/2010 5:30PM – 7:30PM Location: American Folk Museum Type: Museum, Performance
	<ol style="list-style-type: none"> 1. The Grand Manner 	Title: Metropolitan Opera's HD Festival Time: 09/01/2010 8:00PM Location: Lincoln Center Type: Performance
	<ol style="list-style-type: none"> 1. River-to-River Festival 	Title: River-to-River Festival Visiting Governors Island Time: 08/29/2010 10:00AM Location: Governors Island Type: Festival/Fair
	<ol style="list-style-type: none"> 1. Street Fairs 2. Improv 4 Kids! 3. Tony n Tinas Wedding 4. Chicago 5. The Quantum Eye - Magic Deceptions 	Title: Street Fairs Time: 08/21/2010 10:00AM Location: On 6 th Avenue from 42 nd – 56 th Street Type: Festival/Fair

Figure 7. Results on real photo set.

PHOTO COLLECTION SUMMARIZATION

Manually sifting through large collection of personal photos for creating summaries is not only a tedious and inefficient task but also tests human patience. In mobile phones, there are other sets of constraints which makes managing large number of photos a real challenge. First, the screen real estate is limited. Secondly, photos are mostly shot to share with others. Availability of requisite network bandwidth in mobile phones to enable sharing of rich media data is also a big challenge. Hence the need of a system to automatically summarize large photo collections by selecting a subset of the most representative ones. We are actively working on building models, algorithms and evaluation strategies to create optimal summaries. We made significant progress in this area that resulted in several papers [2, 3, 4]. We intend to extend this direction by addressing problems that are specific to mobile phones. In this sections, we motivate the photo summarization problem, propose properties and models to generate effective summaries and define the experimentation framework.

Let us consider the photo corpus of a certain user Joe. Joe shoots lots of photos throughout his life events. It is very difficult to get a quick and representative overview of his photo corpus without manual browsing. Our summarization system automatically generates effective summaries that are extract based subset of the larger corpora. Let us use an example to explain a photo summary. Joe shot almost 5000 photos in 2009. Let us generate a 6 element summary for this corpus. Figure 8 shows a summary generated by random selection of six photos. The location, event type

and time stamp is shown below each photo. From this set, we infer that Joe has been to Beijing, Xi'an and Shanghai in September for sightseeing and to Cambodia in August for a vacation. However, this random summary is not informative about Joe's activities in other parts of the year and only shows the sightseeing or vacation events that Joe participated in. Also the individual photos look really random in nature and may not be very appealing to Joe. Now consider the summary generated by our system shown in Figure 9. This set contains photos spanning from February to December. It also shows the different types of events that Joe participated in cities across the globe. For instance, Joe was in New Delhi for a professional trip while in Irvine for a family birthday event. Most of the immediate family members of Joe (who are important to him) also show up in the photos. Also the photos are arranged in a ranked order. Thus a three element summary will contain the first three photos from this set. Note that, this summary avoids redundancy by being diverse in different spaces (time, location, etc). For instance, the first photo is from China (Beijing). The next photo from China (Xi'an) does not show up until the sixth place in the ranked order. Additionally, the summary has attractive photos and represents important concepts in Joe's life. Increasing the summary size will enable a user to get more interesting information in Joe's life without browsing through the entire photo corpus.



Figure 8: Summary Generated by Random Selection of Six Photos from Joe's 2009 Corpus



Figure 9: An Automated Six Element Summary Generated by Our Algorithm.

Problem Definition

Photo summarization is defined as the process of generating a representative subset of photos from a large personal photo corpus. The corpus may contain photos from a single event (e.g., trip, parties, meetings, etc) or they may contain photos from multiple life events that were shot through months or years. Formally, let the photo corpus \mathbf{P} be a set of N photos, $\mathbf{P} = \{p_1, p_2, p_N\}$. The goal of summarization is to find a set S (with S a subset of \mathbf{P} and $|S| \ll |\mathbf{P}|$), which represents \mathbf{P} in an effective manner.

Thus, photo summarization can be modeled as a subset selection problem. Note that, there are NCM possible summaries of size M for photo corpus of size N , which is exponentially large for any reasonable M and N . Hence it is inefficient to generate and evaluate all possible summaries. However, only a few of these summaries are actually effective. In the following paragraphs we discuss the logic behind choosing the properties that determine effectiveness of summaries, and build the summarization framework based on them.

An effective subset summary should satisfy some desirable properties, which are:

Quality: A photo summary should be interesting to the subject. Quality (*Qual*) of a photo summary determines the aggregate interestingness or attractiveness of the photos present in it. We define the metric *Qual* by integrating various signals like absence of noise, color contrast, presence of portraits and landscapes.

Diversity: Diversity of a summary is a measure of its non-redundancy. A size constrained summary should avoid repetitions and should not contain redundant information. To achieve these goals, the photos in the summary should be diverse. Diversity of the summary S can be modeled as an aggregation of the mutual distances (**Dist**) of the photo pairs. In this research, we use minimum of the pairwise distances of the summary photos as the summary diversity .

Coverage: Coverage ensures that the important concepts present in the corpus are also represented in the size constrained summary. A summary should be a good representative of the larger corpus it is created from. We define the measure *Cov*, which denotes the number of photos in photo corpus \mathbf{P} which are represented by a photo p . Coverage of a summary S , is computed by the aggregating the *Cov* values of every photo in S .

Formulation of the Summarization Problem: We model summarization as a subset selection function $\mathbf{F} : \text{Qual} \times \text{Div} \times \text{Cov} \rightarrow \mathbf{R}$. A good summary is a subset which maximizes \mathbf{F} given the size constraint $|\mathbf{S}| = M$.

$$\mathbf{S}^* = \arg \max_{\mathbf{S}^* \subseteq \mathbf{P}} \mathcal{F}(\text{Qual}(\mathbf{S}^*), \text{Div}(\mathbf{S}^*), \text{Cov}(\mathbf{S}^*, \mathbf{P}))$$

The summarization objective stated in above is a classical multi-objective optimization problem. It can be shown that exact optimization of Div and Cov can be mapped to NP-Hard problems. Hence, we choose approximate optimization. We convert the multi objective function into an single objective function by weighted linear combination. Thus the summarization objective can be reformulated as follows:

The weights can be input by the user, thus generating different summaries as per their needs. In [2, 3] we propose a greedy approach to solve the above optimization function.

Multimodal Information Used to Compute the Summary Properties

Personal photos contain a host of contextual data in addition to the content (pixels). Some of them are captured by various sensors on the camera and some are user or community contributed. A photo p in an corpus is represented by the tuple (x, y) , where x is a set of real valued quantitative attributes and y is a set of discrete categorical attributes or concepts. x is composed of pixel features, time and EXIF-based camera parameters (e.g., exposure time, focal length). The set y contains five concepts: location, event type, visual, temporal and face. The concepts can be generated from the community contributed textual data (e.g., tags, album names, descriptions etc), the image metadata (e.g., GPS induced geotags) or can be predicted using machine learning algorithms on the quantitative attributes. The visual concepts include four different scene types: outdoor day, outdoor night, indoor and sunset. A discrete set of temporal concepts is obtained by clustering the time stamps of the photos in a collection. Each temporal concept may signify a particular event that took place in a user's life. Event types denote a set of popular event categories that are present in consumer photo collections, e.g., birthday, trip, party etc. We leverage on the personal event ontology benchmark proposed by researchers at Kodak, to define these event categories. Location concepts are discrete city names denoting the geographical region where the photo was shot. We use a publicly available geo-database (Geonames.org) and the geotags present in photos to define the location concepts. Face concepts are set of unique faces present in a photo collection. We assume that faces are either manually

tagged (e.g., Facebook's tagging feature) or are predicted by a face recognition system (e.g., Picasa or Iphoto). All this heterogeneous data, along with the pixel feature are used to compute the effective summary properties.

Summarizing multiple photo streams from community events

Multiple people carry their smart phones for various community events. They shoot real time photos of the events and upload them to some photo sharing platform hosted on the cloud. For instance, many people may be shooting photos at a wedding, birthday party, sight seeing or similar personal events. In recent times, platforms have come up (e.g., Color) that allow real time sharing of photos from public events e.g., concerts, sports, games, etc. Sifting through such photo streams from multiple contributors may be tedious. A real time photo summarization algorithm can generate a representative overview of the event by using photos from multiple users.

CONCLUSION

Considering importance of smart phone cameras and visual information, we are developing concepts and techniques to facilitate use of these cameras for various emerging applications. In this paper, we first proposed use of EMPT and then discussed three projects that are building an environment for creation and use of EMPT in various applications. We are unifying these projects into a complete environment that we call Experiential Mobile Media Environment (EMME). We will present experience and results with that in the final paper.

REFERENCES

1. U. Westermann and R. Jain. Toward a common event model for multimedia applications. *Multimedia, IEEE*, 14(1), 2007.
2. Pinaki Sinha, Sharad Mehrotra, Ramesh Jain. Summarization of Personal Photologs Using Multidimensional Content and Context. *ACM Intl Conference on Multimedia Retrieval (ICMR) 2011*.
3. Pinaki Sinha, Ramesh Jain. Extractive Summarization of Personal Photos From Life Events. *IEEE International Conference Multimedia Expo (ICME) 2011*.
4. Pinaki Sinha, Sharad Mehrotra, Ramesh Jain. Effective Summarization of Large Collections of Personal Photos. *International World Wide Web Conference (WWW) 2011*.
5. Smith, J.R., Chang, S.F. Large-scale concept ontology for multi-media, *IEEE Multimedia*, 13(3), 2006.
6. Gruber T., Liu, L., Tamer Özsu, M. In the *Encyclopedia of Database Systems*, Springer-Verlag, 2009.
7. Smeulders, A., et al., Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.