Taking the Tags With You: Digital Photograph Provenance

Brandeis Marshall
Purdue University
Computer and Information Technology
401 N. Grant Street
West Lafayette, IN 47906
brandeis@purdue.edu

Abstract—Prior to digital photography, the back of the glossy photo paper served as the canvas to write the important photo details. The photos were then transferred to a series of photo albums. The general public creates too many pictures to sustain the physical photo album. One byproduct of taking digital photos is the automatic number scheme embedded in the digital device. The semantic context of any photo can not be represented with this automatic labeling method. Hence, the user is responsible for concisely renaming the digital images. The rise of digital pictures creates a need for more reliant on storage technology for longevity. We present a suite of annotation tags that capture provenance details. We classify consumer photography provenance in terms of photo manipulation, photo clustering and photo semantics.

Keywords—image retrieval, data lineage and provenance, photo properties, XML

I. INTRODUCTION

Retrieving images or photos have generally taken one of two approaches: content-based image retrieval (CBIR) [1], [2] or image annotation or tagging [3]. In CBIR, the low-level features, such as color and shape, are computationally processed to reveal the high-level semantics and determine the relevancy of an image. Image annotations use humans to label images using a concise set of keywords. The pervasiveness of digital and phone cameras have created a sharp growth in digital photos. In turn, the reliance on humans to label photographs increases since annotating is faster than the CBIR approach [4].

Digital photos have their own challenges. The digital camera allows users to take multiple photographs of the same or similar scenery. The tedious job of managing photographs after downloading them from the camera onto a computer falls on the consumer [5]. Digital cameras typically auto-label photos according to a numerical scale while users rename them based on a date, place and/or event. Users then struggle to organize their photos. A typical photo download session usually consists of multiple images over a duration of time. A popular trend in sharing user photo collections includes online photo sharing applications like Flickr1 and Picasa2. These applications allow for single and group photo tagging. However, the modular platform of these applications limit its portability and extensibility: (1) users have a username and password, (2) users have restricted storage capabilities and (3) photo annotations are non-transferable.

Previous studies [3], [6] focus on the importance of social relationships when annotating photographs. The motivation for tagging photographs is strongly driven by the social circle of an individual i.e. family and friends. People want to share their memories between family and friends and in turn annotate their photographs as a medium of communication to provide the extra contextual information. The online photo sharing portals are not designed for long-term photo storage, which is becoming an increasing need for users. Since personal computers contain abundant storage access, users have the opportunity to exploit their hardware for multimedia data storage.

We address in this paper the problem of archiving the annotated content from digital photographs. We extend our social networking graph (SNG) model [7] to include data lineage. Our SNG model categories photographs by connecting the photo’s objects and the corresponding relationships amongst them. We identify each photograph in our framework with five attributes of objects, place, occasion, time and associations. We show how our SNG model can be represented using XML and then present a suite of annotation tags that capture provenance details. We classify digital photography provenance in terms of photo manipulation (modification), photo clustering (grouping) and photo semantics (context). We extend the flexibility of our social networking framework by designing photo clustering and semantics tags to consider object, place, occasion, time and association metadata. As an image testbed, we select the MIRFLICKR-25000 [8] image collection due to its diversity and size. We report the initial statistics related to the MIRFLICKR testbed.

II. RELATED WORK

The preservation of data has been a widely explored research area [9], [10], [11], [12]. Moreau et al. [9] outline the provenance lifecycle of creating, recording, querying and managing that must exist for any provenance-aware application. The authors stress that documentation must be structured and flexible in order to prevent system bottlenecks. The discussion of provenance in cloud computing [10] has begun due to...
the limited design of cloud storage services for metadata. Reddy and Seltzer describe the provenance requirements over
the cloud as coordination between storage and computing facilities, allowing customers to record provenance of their
objects, provenance data consistency, long-term persistence, exposing provenance to external use and provenance security.
A majority of these requirements can be applied for any provenance-aware application.

Boye and Frew [11] conducted an extensive survey of data lineage retrieval in the area of scientific data processing. Data
lineage or provenance has two well-known direction: backward links to ancestry content and forward links to descendent
contents. The majority of the prior provenance research has focused on the scientific domain and its phases of experiment
design, experiment execution and data analysis. In general, each experiment iteration is its own silo of data and relating
experiments and analysis is performed at a later time. Our SNG provenance model does not contain a data analysis
experiments and analysis is performed at a later time. Our SNG provenance model does not contain a data analysis

A majority of these requirements can be applied for any
provenance-aware application.

III. SNG Framework

The management, browse and retrieval of digital pho-
tographs on a personal computer is a hierarchical collection
of semi-related directories. The date and event name are
used consistently to organize photos [5]. However, users must
manually create and rename folders to make future viewing
or retrieval accessible. The social networking framework can
capture relationships amongst photos based on objects within
the photos as well as metadata associated with a group of
photos.

A. Photo Representation

We describe the content of a photo using who, how,
where, what and when attributes of a photograph. We classify
these five major attributes as object, association, place, event
and time respectively. A photo can be described using a
combination of any of these five attributes as shown below:

We define a photo (Photo) as a 5-tuple:

\[ \text{Photo}(t, o, p, e, a) \] (1)

where timestamp \( t \) is not null. In older digital photographs, the date
of a photo may not be available. If a photo does not include a
date, then we assign a default timestamp. These five attributes
capture human-centered properties that are typically searched.

These attributes are further described below:

1) Timestamp \( t \). A digital camera image contains the date of “when” the photograph was taken.

2) Object \( o \). An object is any entity “who” or “thing” that

3) Place \( p \). The location or landmark “where” attribute that

4) Event \( e \). Any image may describe a “what” event or occasion such as wedding, birthday or vacation.

5) Association \( a \). An association denotes the “how” attribute of the objects. This relationship may be personal

Using the five attributes defined in equation (1), a photo-

graph is represented as an undirected social networking graph

\[ \text{SNG}(V, E) \]

where \( V \) denotes the set of all objects and \( E \)
denotes the set of all relationships between any two objects in
the graph. Based on the 5-tuple, the date, place, event and
association content related to any particular photograph
is stored as an edge \( \text{edge}(o_i, o_j) \) for a given pair of objects
\( o_i, o_j \).

Since objects may appear in multiple photos, SNG can have
multiple edges between the objects \( o_i \) and \( o_j \) of the form

\[ \text{edge}(o_i, o_j) \]

where \( \chi = \{t, p, e, a\} \). A photograph \( P \) can be
transformed to a clique, which constructs edges between every
object pair. Hence, the social network graph is a collection of
subgraphs for ease in browsing and retrieval as represented in
equation 2.

\[ \text{subgraph } G(\text{Photo}) = \text{Clique}(V', E') \text{ s.t. } \]
\[ V' = \{o_1, ..., o_n\} \text{ and } E' = \{edge_1, ..., edge_m\} \] (3)

Our framework is built for flexibility and extensibility by only
requiring time to serve as the necessary attribute. The objects
can be added or updated as different users view the photo-

graph. The place, event and any associations that interact with
the objects can be added or modified as time permits. Thus, our
framework lends itself to semi-automatic labeling techniques
since commonly searched components are centralized. The
semi-automatic labeling of event and association can occur by
leveraging the existing date and object content, respectively. Thus, repeatedly annotating photos with the same tag could be significantly reduced. The work of semi-automated labeling for photos is part of future work.

Fig. 1. Social Network Graph representing the 3 photographs: Photo\textsuperscript{0} refers to the place the photo was taken, Photo\textsuperscript{1} refers to photo’s event and Photo\textsuperscript{1} refers to a photo’s date

In the case of multiple photos Photo\textsubscript{1}, Photo\textsubscript{2} and Photo\textsubscript{3}, we can view these photos components in a network graph such as Figure 1. We have a series of three photos with 4 objects represented in each photo. The notation Photo\textsubscript{y} refers to the photograph containing attribute \( x \)'s information. Photo\textsubscript{y} = Hawaii refers to the second photograph being taken at a place in Hawaii. Photo\textsubscript{y} = home refers to the first photograph being taken at place denotes as home. Photo\textsubscript{y} = vacation refers to the first photograph occurring during the event of a vacation. Photo\textsubscript{y} = Photo\textsubscript{y} = July 2008 refers to first photograph being taken in July 2008. Photo\textsubscript{y} = Aug 2008 refers to the third photograph being taken in August 2008.

Photo\textsubscript{1} and Photo\textsubscript{2} have disjoint objects; however the objects of Sky and John may be related with respect to time, place or event. Photo\textsubscript{2} and Photo\textsubscript{3} include one common object, John. In a search of “pictures containing John”, Photo\textsubscript{2} and Photo\textsubscript{3} would be returned whereas “pictures associated with John” would retrieve Photo\textsubscript{1}, Photo\textsubscript{2} and Photo\textsubscript{3}.

\begin{verbatim}<pic>
<name> filename </name>
<time>manual or automatic timestamp </time>
<object> object’s name </object>
<object> object’s name </object>
<place> location name </place>
<event> event name </event>
<association> relationship label </association>
<source> \( a_1 \) </source> sink = \( a_2 \)
<source> \( a_1 \) </source> sink = \( a_2 \)
<edge>
<source> \( a_1 \) </source> sink = \( a_2 \)
<source> \( a_1 \) </source> sink = \( a_2 \)
<edge>
<edge>
<edge>
<edge>
<edge>
<edge>
...</verbatim>

**TABLE I**

<table>
<thead>
<tr>
<th>SNG XML REPRESENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

In order to effectively implement our SNG framework, we make use of the XML, primarily due to its semi-structured format and easy database extensibility. In Table I, we show a snippet of the XML for any given photograph. Each photograph is encapsulated by the pic tag with single values for time, place, event and association tags. The object and edge tags may be repeated for the multiple instances. Even though SNG edges are undirected, our XML forces a directed edge of source and sink object. However, the reverse edge is implied in the database management system. We extend the SNG XML representation to relate photographs for preserving data lineage.

**IV. DIGITAL PHOTOGRAPHY PROVENANCE**

Prior to digital photography, the back of the glossy photo paper served as the canvas to write the important details of the photograph. The photos were then transferred in random or orderly manner into a series of photo albums. As people move away from physical photo albums, consumers become more reliant on storage technology for longevity. To effective archival and browse digital photography, the photowork of reviewing, downloading, organizing, editing, sorting and filing [5] must expand to also include provenance. Our approach to photo provenance considers extending the SNG XML document to capture the relationships important to consumers.

We consider three main categories of photograph provenance that will be added to the social networking framework discussed in Section 3. We describe our modification, grouping and context categories in detail below. Generally, the reviewing, downloading and editing are part of the modification category, organizing and sorting are part of the grouping category and the anecdotal details are stored within the context category.

- **Modification**: Any digital camera photo can be altered through cropping, color and size adjustment and/or image rotation.
- **Grouping**: Any photo collection can contain many overlapping clusters based on related objects, places, times, events or associations.
- **Context**: Each photograph has a story. Thus, we include this category for extended notes and commentary related to the given photograph.

Table II displays the complete SNG XML representation with data lineage included for a single photograph. We expand our SNG XML representation to allow the mod, group and cnote tags, which correspond to the modification, grouping and context categories, respectively. We embed the 5-tuple attributes within the group tag in order to promote information reuse. The edge information is stored at the end of the pic tag. We assume the data associated with any of these tags to be handled manually. In the future, we plan to provide a semi-automated algorithm that would reduce the user workload and minimize human error. We discuss each category’s tags and provide an SNG XML provenance example.

**Modification.** A fundamental component of digital photography is the ability to manipulate the original raw image. We develop a modification tag, \(< \textit{mod} > \ldots < /\textit{mod} >\), to indicate that the current image has been altered.

Traditionally, a user may minimally change a single raw image to remove blurriness or red eye. In the flat file infrastructure of operating systems, the connection between two
files are made by creating a folder. We reflect the parent-child relationship between a modified photo to the original image through an internal tag, <parent>...</parent>.

Within our modification tag, we record the selected types of alteration as an internal tag. We use a crop tag, <crop>...</crop>, that lists the percent of the original retracted. The resize tag, <resize>...</resize>, stores the new height and width dimensions. Lastly, the color adjustment tag, <coloradj>...</coloradj>, logs the changes to the contrast, hue and/or saturation of the image.

The example below (Figure 2) displays how our framework would represent the parent-child relationship. Typically, the user imposes an ordered naming scheme such as im49.jpg, im49v1.jpg, im49v2.jpg as a method to easily re-find similar photographs. Our approach reduces the need for this user-defined naming convention. The photo retrieval process also becomes less computationally costly with the ability to explore the parent-child relationships using the XML structure.

**Grouping.** As digital camera users, we take photographs in bursts. However, the digital camera records the photo are independent entities. Thus, part of photowork encompasses categorization of digital images into folders and directories. However, one photo may be related to more than one folder. Consumers tend to store each image in only one folder in order to reduce image redundancy on their computing system.

We design a tag, <group>...</group>, that captures some of the multiple dimensionality of a single photo. Internal tags covers the photo attributes: <object>...</object>, <time>...</time>, <place>...</place>, <occasion>...</occasion> and <association>...</association>. For example given a set of photos that contain place attributes of ‘Bordeaux’, ‘Nice’, ‘Paris’, ‘Louve’ and ‘Eiffel Tower’, the place internal tag within <group>...

... < /group > can be labeled appropriately as ‘France’. Now, a consumer could more effectively search using the more inclusive group’s place tag ‘France’ keyword in lieu of the individual place attributes.

Figure 3 shows an example usage of the group tag. The original 5-tuple SNG attributes are encased in this tag to allow for ease in labeling propagation. If a series of photographs should be group tagged, the time, place, event and association data can be repeated for each photograph since the values do not change. However, the objects within the photograph may vary in which case the user can modify the object tags individually. When browsing or retrieving certain information, the photographs with same or similar labels can be located easily.

**Context.** One byproduct of taking digital photographs is the automatic number scheme embedded in the digital camera. The semantic context of any photo can not be represented with this automatic labeling method. Hence, the consumer is responsible for concisely renaming the digital images. An image’s filename can not hold a complete description for photo sharing or keepsake purposes. We design an unstructured text tag for commentary notes called <cnote>...</cnote>. We also use the <cnote> tag to interrelate our photo attributes. For instance, a <cnote> of “Jane’s September 2006 wedding in Scotia, NY” could be cross-listed as <group> internal tags: ‘Jane’ (object), ‘September 2006’ (time), ‘wedding’ (event) and ‘Scotia, NY’ (place).

In Fig 4, we show how the cnote tag can provide a semantic description of a photograph. This semantic description allows users to capture the narrative (“a picture is worth a thousand words”) that can not be articulated in a handful of keywords. The contents within the cnote tag can serve as the photograph’s caption for photo sharing. This free form text provides associations can also be used in improving image retrieval effectiveness.

**V. Future Work**

We will be extending our social networking graph model [7] to incorporate our proposed provenance model. We have
downloaded the MIRFLICKR-25000\(^3\) collection consists of over 25,000 images in 24 categories with a high interestingness rating. We have initially processed 13 categories with the most popular tags. The popular tags includes animals, bird, clouds, dog, flower, lake, night, river, sea, sky, sunset, tree and water. In Table III, we display our initial categories indicating the number of photos with and without tags and the average number of tags for a given photo. We observe 9,675 images with 16,868 unique tags. Some of the image tags containing special characters (e.g. México) and non-english characters, which we ignore. We also notice that some images are labeled with the date or time information such as the year attribute ‘2008’.

We are currently adding the mod, group and cnote tags to create our SNG XML provenance representation. We will then be manually adding content to the mod and cnote tags using computing student volunteers for our experimental evaluation. The experimental evaluation will compare the effects of increased file storage to the effectiveness of image retrieval through precision and recall measures and scalability through response time.

**REFERENCES**


\(^3\)http://www.flickr.com/photos/tags/:Retrieved on October 2009


