A COLLECTION-ORIENTED METADATA FRAMEWORK FOR DIGITAL IMAGES

William Ku, Mohan S. Kankanhalli* and Joo-Hwee Lim#*

School of Computing, National University of Singapore* / Institute of Infocomm Research, Singapore#
{kucheech, mohan}@comp.nus.edu.sg / joohwee@i2r.a-star.edu.sg

ABSTRACT

A digital photo can "tell a thousand words" through the use of its metadata and as it is usually part of a collection, metadata management, reuse, propagation & inference could be achieved via its association with a collection. However, there is not much work on metadata management, reuse, propagation & inference, particularly on a group basis. In this paper, we proposed a collection-oriented metadata framework which provides a basis for metadata management, reuse, propagation & inference and demonstrated the utility of such a framework.

1. INTRODUCTION

The advent of digital imaging technologies has empowered one to take a lot more photos with a digital still camera (DSC) than one would have taken with an analog camera, cheaply and effortlessly. As the saying goes: "A picture tells a thousand words". To tell these thousand words for a photo, in an objective manner, one would require the presence of the photo taker or at least someone knowledgeable on the content and context of the photo. The digital image offers a significant advantage over its conventional counterpart: its capability to house these thousand words and more. When and where the photo was taken, under what kinds of lighting conditions was it taken, the human subjects and landmarks in the photo, the purpose of the photo. All the answers to these questions and more could be encoded, in the form of metadata. Indeed, there are many digital image metadata specifications that could help to tell the story behind a photo, such as EXIF [4] (image capture device settings at point of image capture).

These specifications allow for manual textual annotations such as captions and (narrative) descriptions, and even audio annotations. While these specifications may be extensible to cater for new terms, they are essentially "fill-in-the-blanks" schemas that cater to specific application domains and require either manual annotation (which is usually a tedious, inconsistent and erroneous process) or automatic annotation which unfortunately produces unsatisfactory results presently due to the semantic and sensory gaps [1, 9]. Furthermore, these schemas do not provide specifications for metadata manipulation operations such as metadata propagation/reuse and inference. We are concerned with having good metadata because it can be used to provide (but not limited to) (1) administrative information, (2) content description, (3) for context, and (4) for search and retrieval.

While an image can exist as an individual entity, it is usually part of an image collection such as a photo album or an online photo collection. Even within a DSC, the images stored in the memory card can be regarded as part of a collection. In certain DSCs, it might even be possible to sort images into folders. Metadata management, reuse and inference could be achieved via association with an image collection. We shall illustrate this with the following examples.

- *Common metadata*: One would usually take pictures with just one DSC. Hence, the EXIF metadata contained within these photos would contain the same camera settings with some exceptions such as focal length and whether the flash was used. Thus within an image collection, it would be practical from the metadata management perspective, to extract these common metadata. Care has to be taken to reinstate these common metadata should one make a copy of an image for use beyond the domain of the image collection, for example photo-sharing via email.
- *Metadata reuse*: One might like to provide some common information for the photos, for example administrative information such as ownership, access and usage rights. Thus by providing these common information at the group level, new photos entering the collection will inherit these common metadata.
- *Metadata propagation and inference*: Suppose one has an image collection pertaining to a birthday party. A friend emailed him some photos taken during this birthday party, complete with several annotations. Thus, when the new photos are inserted into the image collection, some form of metadata propagation and inference can take place. For example, the name of a friend from the new photos may get propagated.

In this paper, we propose a metadata framework that provides for metadata reuse, propagation and inference on a photo collection basis.

2. RELATED WORKS

There are some work on metadata propagation, reuse and inference on a group basis. [7] described LOCALE which tags unlabeled photographs using shared information based on other photos taken in the same area. [3 expounded the social-temporal-social context influencing the metadata values of images taken by a group of cameraphone users. [2, 5] described Snap2Tell which matches a photo taken with cameraphone with a database, using content-based features and metadata. [8] put name labels on photos, using event and location groupings to suggest name labels. [6] proposed a novel automatic mechanism for XML based video metadata editing, catering for conflict resolution and regularization operations.

3. PROPOSED METADATA FRAMEWORK

We observed in our literature survey that there is no “one-size-fit-all” digital image metadata schema but instead there are several different types of schemas that provide for specific domains and applications. It is possible to assemble any customized schema by using some (or parts of) these schemas, assuming semantic, syntactic and structural interoperability as defined by [11]. For example, one could specifically choose a schema that provides for the technical attributes of the image device (for instance EXIF) and another schema that provides for feature-based content description (such as MPEG-7 [10]). However, these schemas do not provide group management. Thus, we would like to propose a framework that allows for such modularity while at the same time, provides for metadata management and operations at the group level. Here we assume that metadata semantic, syntactic and structural interoperability are in place.

3.1. Basic Framework Formulation

In this section, we outline our basic metadata framework formulation and whenever possible, provide illustrations.

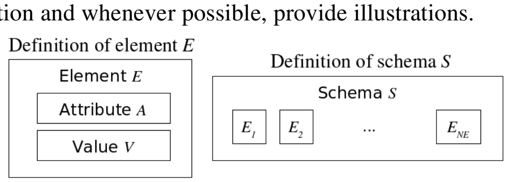

Fig. 1: Basic units of metadata

3.1.1 Element definitions

A metadata element E is defined as an ordered pair *(A,V)* where A is an *attribute* and V is a *value*. An attribute is a textual descriptor specifying a particular property or characteristic while the corresponding value is an instance of the attribute and pertains to a particular data type determined by the nature of the attribute. We illustrate some fundamental element concepts in Fig. 2.

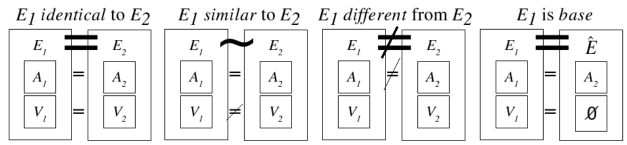

Fig. 2: Some fundamental element concepts

E is defined to be *unique* in a schema S when there are no other elements *identical* to it in S. There shall be no two or more identical elements within a schema. Similar elements can exist within a schema and shall be considered as unique elements. However, the use of similar elements has to be handled with care so as to avoid ambiguous context. E is defined to be *base* (denoted as $\hat{E}$) when its value is not instantiated or set to a default value of its data type, denoted as $\varnothing$. All elements are *similar* to their base element counterparts : $E \sim \hat{E}$. E can be reduced to its base element : $E \rightarrow E = \varnothing \Rightarrow E = \hat{E}$.

3.1.2 Schema definitions

As illustrated in Fig. 1, A schema S is a set of unique elements, $S = \{E_1, E_2, ..., E_{NE}\}$, where NE is the number of unique elements.

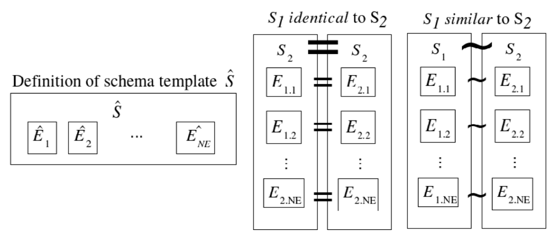

Fig. 3: Some fundamental schema concepts

S is defined to be *base* (also defined as a *schema template* $\hat{S}$) when every of its elements is *base* (see Fig. 3). There shall be no identical or similar elements within $\hat{S}$. S_1 is defined to be *empty* when $|S_1| = 0$ but for $\hat{S}_2$, $|\hat{S}_2| \neq 0$. S_1 is defined to be *similar* to S_2 when $\hat{S}_1$ and $\hat{S}_2$ are identical. S_1 is defined to be be a *subset* of S_2 when every element in S_1 is similar to at least one element in S_2 or alternatively when $\hat{S}_1$ is a proper subset of $\hat{S}_2$. S_1 is defined to be *different* from S_2 when S_1 is a subset of S_2 or at least one element in S_1 is not found in S_2.

3.1.3 Image metadata structure definitions

We define the image metadata structure IM of image I as $IM = \{S_1, S_2, ..., S_{NS}\}$, where NS is the number of unique image metadata schemas : $S_i \not\subseteq S_j , \forall i, j \in \{1, \ldots, NS\} \wedge i \neq j$. Some metadata redundancy may be allowed but there can be no identical or similar schemas within IM.

$$S_i \cap S_j \neq \varnothing , S_i \neq S_j , S_i \simeq S_j \forall i , j \in \{1, \ldots, NS\} \wedge i \neq j$$

IM_1 is identical to IM_2 when

$$IM_1 = IM_2 \; iff \; |IM_1| = |IM_2| = NS \wedge IM_1 . S_i = IM_2 . S_i , \\ \forall i \in \{1, \ldots, NS\}$$

IM_1 is similar to IM_2 when

$$IM_1 \sim IM_2 \; iff \; |IM_1| = |IM_2| = NS \wedge IM_1.S_i \sim IM_2.S_i, \quad \forall i \in \{1, \ldots, NS\}$$

IM_1 is a subset of IM_2 when every schema in IM_1 is similar to at least one schema in IM_2 :

$$IM_1 \subset IM_2 \; iff \; \exists j \in \{1, \ldots, |IM_2|\} \; s.t. \; IM_1.S_i \simeq IM_2.S_j \, \forall i \in \{1, \ldots, |IM_1|\}.$$

3.1.4 Group metadata structure definitions

We define a image collection G to be $G = \{I_1, I_2, \ldots, I_{NI}\}$, where NI is the number of unique images : $I_i \neq I_j, \forall i, j \in (1, \ldots, NI) \wedge i \neq j$ and its corresponding group metadata structure to be GM where $GM = \{S_1, S_2, \ldots, S_{NGS}\}$, where NGS is the number of group metadata schemas. The properties that have been defined for IM also apply to GM. In addition, there shall be no schemas in GM that are identical or similar to those in the respective image metadata structure IMs of the images in G :

$$IM_i.S_j \notin GM, \forall i \in \{1, \ldots, |GM|\}, j \in \{1, \ldots, |IM_i|\}$$

The images in a collection G may not have the same image metadata structure although it would be recommended to format all the images to have the same image metadata structure IM. We would assume that this is the case here.

3.2. Metadata Conflict Resolution and Inference

Conflict Resolution is an essential component in metadata propagation because it provides the basis for the update or integration of the metadata value of a metadata element from potentially multiple sources (see Fig. 4 for overview). In conflict resolution, we are concerned with the reconciliation of two differing values of the same metadata element in question while metadata inference has a much wider scope. We shall examine conflict resolution in greater details first and discuss metadata inference later.

3.2.1 Metadata Conflict Resolution

In the base case, we have two instances of a particular image. These two instances have metadata schemas S_1 and S_2, both of which have the same metadata element E. Thus, $S_1.E$ and $S_2.E$ have the same metadata attribute A with respective metadata values of V_1 and V_2. We define metadata conflict resolution as the reconciliation process involving V_1 and V_2 such that one gets *updated* by the other or both of them are *integrated* to render a third value. We formulate these two operations as follows.

Update: V_1 is to be superseded by V_2. Here V_1 is an obsolete or uninitialized value that is to be updated by a more recent and correct V_2. For example, we may have a metadata element describing the event-type for the image which is not annotated at the point of image and it is to be updated through the introduction of another (better annotated) image. The update operation can be extended to provide for update of more than two metadata values by reducing it to the base case involving two metadata values and handling them one pair at a time.

Integration: V_1 and V_2 is to be integrated using some operations (such as addition or multiplication) to generate V_3 which will then supersede both V_1 and V_2. Take for example a metadata element that captures the maximum number of daily user views for the image and we have conflicting values for V_1 and V_2. In this case, we want to take the larger of these two values. The integration operation can be extended to provide for integration of more than two metadata values. In fact, it could be generalized to any logical functions. Technically, the update operation is a subclass of the integration operation. However for the purpose of our discussion, we separate them.

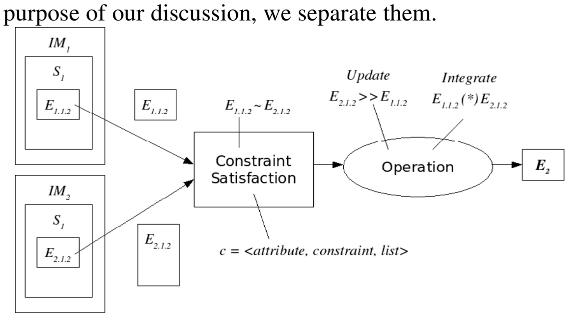

Fig. 4: Overview of metadata conflict resolution

3.2.2 Constraints

Constraints are essentially mandatory conditions which must be satisfied in order for the update and integration operations to be permissible. We formulate the notion constraints to be be an tuple c = <*attribute, constraint, list*> where *attribute* refers to the metadata attribute and *constraint* is the name of a specific unique constraint while list denotes the list of acceptable or compatible parameters for the constraint.

3.2.3 Metadata Inference

Metadata inference is similar to (and technically is a superset of) metadata conflict resolution. Given that we are to determine the value V of a metadata element E in metadata schema S associated with an image I (i.e. in image metadata structure IM or group metadata structure GM), metadata inference to determine V can take one of the following forms: (1) One or more metadata elements (but not E) in S, (2) Metadata element(s) from one or more schemas (but not S) associated with I, (3) Metadata elements(s) from schema(s) associated with one or more images (but not I), (4) Other information sources such as digital calender or webpages, and (5) Any combination of the above!

As with metadata conflict resolution, the update and integration operations are applicable to metadata inference.

4. IMPLEMENTATION/EXPERIMENTAL RESULTS

We made use of some photos (available at [12], see Fig. 5) taken at a panel discussion during the recent ACM Multimedia 2005. In this collection, seven photos (Set A) were taken with a Fuji DSC and contained EXIF metadata (without user comment) while two photos (Set B) taken with a Canon DSC contained EXIF metadata (with user comment) and another two photos (Set C) taken with a cameraphone, did not contain any metadata. In our scenario, Set A photos formed the original collection while Sets B and C photos were contributed photos, added to the collection separately.

Fig. 5: a-3.jpg (left), b-1.jpg (centre) & c-2.jpg (right)

Common metadata: Based on our proposed framework in Section 3, we established that in Set A, out of the 64 available EXIF metadata attributes of these photos, six attributes (namely *Image Created, ISO Speed Rating, Brightness, Bytes of JPEG Data, Image Generated & Image Digitized*) have different respective values (example shown in Fig. 6). Furthermore, the attributes *Image Created*, *Image Generated* & *Image Digitized* are actually duplicates. This was also the case with Set B. Thus within Sets A & B, much metadata could be "exported" as **common metadata** at the group level as shown in Fig. 7.

```
Image Created: 2005:11:10 14:00:08         | Image Created: 2005:11:10 14:00:37
ISO Speed Rating: 200                      | ISO Speed Rating: 160
Brightness: -0.3 EV                        | Brightness: 1.1 EV
Bytes of JPEG Data: 9005                   | Bytes of JPEG Data: 9239
Image Generated: 2005:11:10 14:00:08       | Image Generated: 2005:11:10 14:00:37
Image Digitized: 2005:11:10 14:00:08       | Image Digitized: 2005:11:10 14:00:37
```

Fig. 6: Screenshot of similar attributes between a-1.jpg & a-2.jpg

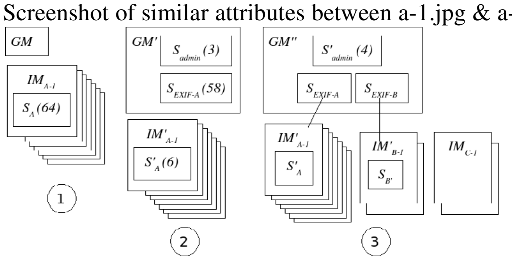

Fig. 7: Grouping common metadata

Since the EXIF metadata of Sets A & B photos are of different versions (we identified 17 identical attributes, 32 similar attributes & 41 different attributes between them) and that camera manufacturers have their own way of handling EXIF metadata, we have two separate sets of common EXIF metadata (see Fig. 7).

Metadata Reuse: Another difference between Set A and Set B photos is that Set B photos contain a *user comment tag*, which can be set to provide a caption. In this case, this has been tagged "Panel Discussion at ACM Multimedia 05" for both Set B photos. In our scenario, we have identified this tag as an important tag and have automatically incorporated it into our S_{admin} schema which also consists the *owner*, *date*, *location* tags. The value of the *date* tag is derived from the EXIF metadata. Here, Set A photos can **reuse** this caption. Set C photos do not have any metadata and hence when they join the group, they can **reuse** at least the S_{admin} metadata. Thus, when Set C photos are shared, at least some metadata is available as compared to none previously. The metadata reused is predetermined and domain-specific.

Metadata inference & propagation: The metadata used for inference & propagation is not restricted to EXIF. We are presently working on possible times for set C photos using visual similarity (see Fig. 8), based on a previous work [2].

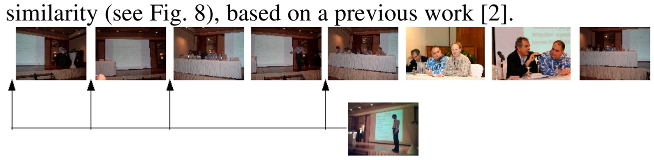

Fig. 8: Possible points of insertion

5.CONCLUSION

We have proposed a collection-oriented metadata framework for digital images that provides a basis for metadata management, reuse, propagation and inference.

6. REFERENCES

[1] Dick C.A. Bulterman. Is it Time for a Moratorium on Metadata? IEEE Multimedia, Vol. 11 Issue 4, pp. 11 – 17, 2004.

[2] Jun Li, Joo-Hwee Lim and Qi Tian. Automatic Summarization for Personal Digital Photos, Proc. of PCM 2003, vol. 3, pp. 1536-1540.

[3] Marc Davis, Simon King, Nathan Good and Risto Sarvas. From context to content: leveraging context to infer media metadata, Proc. of ACM Multimedia 2004, pp. 188 – 195.

[4] Exchangeable Image File Format (EXIF), Version 2.2, JEITA. http://www.exif.org/Exif2-2.PDF, 2002.

[5] Joo-Hwee Lim, Jean-Pierre Chevallet and Sihem Nouarah Merah. SnapToTell: ubiquitous information access from camera. A picture-driven tourist information directory service, Proc. of MobileHCI 2004, pp. 21-27.

[6] Chitra L. Madhwacharyula, Mohan S. Kankanhalli and Philippe Mulhem 2004. Content Based Editing of Semantic Video Metadata, Proc. of ICME 2004, pp. 33-36.

[7] Mor Naaman, Andreas Paepcke and Hector Garcia-Molina. From Where to What: Metadata Sharing for Digital Photographs with Geographic Coordinates, Proc. Of CoopIS 2003, pp. 196-217.

[8] Mor Naaman, Ron B. Yeh, Hector Garcia-Molina and Andreas Paepcke. Leveraging Context to Resolve Identity in Photo Albums, Proc. of JCDL 2005, pp. 178 – 187.

[9] Frank Nack. All Content Counts: The Future in Digital Media Computing is Meta, IEEE MultiMedia, vol. 7, no 3, pp. 10-13, 2000.

[10] MPEG-7 Overview.http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[11] Arlene G. Taylor 2004. Organization of information. Westport, Conn. : Libraries Unlimited, 2004, 2nd edition. ISBN: 1563089696.

[12] http://www.comp.nus.edu.sg/~kucheech/icme06/