

# A Personal Photo Album with Semi-Automatic Metadata Generation

Supheakmungkol SARIN\*    Toshinori NAGAHASHI†    Tadashi MIYOSAWA†  
and Wataru KAMEYAMA\*

\*GITS, WASEDA University  
mungkol@fuji.waseda.jp  
wataru@waseda.jp

†Corporate R & D Div., SEIKO EPSON Corporation  
nagahashi.toshinori@exc.epson.co.jp  
miyosawa.tadashi@exc.epson.co.jp

## 1 Introduction

Digital cameras have changed the world in many ways. Today people can take hundreds of digital photos for any event. However, with this incredible growth of digital photos, there is still a lack of effective management software that helps organizing and finding photos in an efficient and accurate manner.

Presently there are two major techniques that have been proposed and used. The first one is the Content-Based Image Retrieval (CBIR), which can be computerized to extract perceptual information like colors, shapes, and textures from the photos automatically. However there is an absolute limit in content-based technology. Some information cannot be interpreted computationally especially when dealing with abstract information inside the photos. The other method uses the non-pixel information of the photos: contextual information that related to the photos as keywords to identify the photo. The information is referred to as metadata. This method can reduce semantic gap but it is laborious, time consuming and costly especially when applying to huge collection of photos.

Lately, Global Positioning System (GPS) technology made it feasible to add timestamp and location information to digital photos, namely the exact coordinates where each photo was captured. Although, location-aware cameras are not widely available at the time being, the GPS-enable mobile phones are available in many countries including Japan.

Given the current technology and that not many applications set to profit from it, in this paper we envision utilizing the geo-reference information in conjunction with time information as key tools to generate semi-automatic metadata for photos.

## 2 Related Work

Various researches on how to annotate photos using metadata have been going on actively in the last 10 years. Today's popular commercial photo album applications also attempt to allow efficient annotating photos. For example, Photoshop Album divides label classifications into categories such as place, event and people. There is also a trend toward collaborative annotation. People can upload their photos to a website. Photos can be labeled by tags or captions, and tags can be assigned by any user, not only the owners of a particular photo. An example of this community labeling is Flickr [1].

Using GPS coordinates with digital photos is fairly recent approach. A relatively close work to ours is that of M. Naaman et al. who try to utilize geo-referenced information to organize collections of photos [2].

## 3 Photo Album with Semi-Automatic Metadata Generation

### 3.1 System Features

Our system aims to provide a comprehensive set of facilities for the management of a digital photo collection. Therefore, all the basic functionalities of the personal photo album such browsing, searching, organizing are included in the system. Moreover we integrate a special feature to generate metadata for photos. This means that keywords of each photo will be assigned to the photos automatically by our system and users will be asked to select the appropriate candidates proposed by the system. The keyword will be used in addition to the photos as a key for auto-indexing. Thus provides efficient and accurate searching and browsing result. The metadata generation process is described in the next section.

\*GITS - Graduate School of Global Information and Telecommunication Studies

### 3.2 Generating Contextual Photo Metadata Semi-Automatically

We assume that the digital camera is equipped with GPS receiver. Our proposed model to semi-automatically generate photo metadata can be illustrated as in Figure 1. First of all, photos are analyzed to extract embedded EXIF [3] data which contains all the information about camera settings and scene information when they were taken (step 1). Timestamp and the coordinate set (longitude and latitude) are part of the extracted information. Then we perform a translation of the pair coordinates of each photo into location names such as cities or land marks by trying to locate the nearest and the most well-known referential location to the photos.

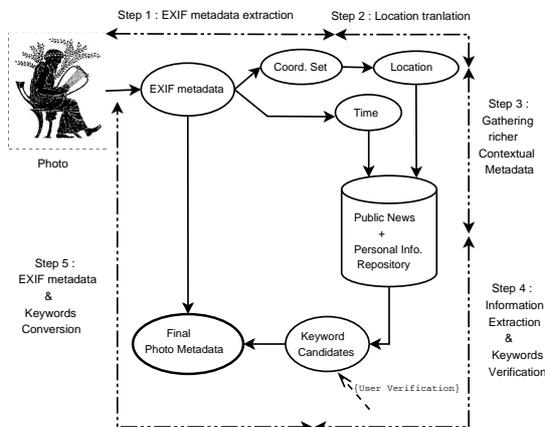


Figure 1 : Processing steps in semi-automatic photo metadata generation

From the fact that when searching for a photo, usually users may not remember exactly what the photo is all about. In stead, users may try to associate some other information to the photo. For instance, a user may remember that the photo was taken in the period when the Summer Olympic Games was held in Tokyo. Moreover, a user may remember that at the day the photo was taken it was also a day when he or she had a meeting with a representative from an X company. This shows us clearly that the key information users may associate to the photo comes from public information like news and their personal information such as their schedules. Therefore in our system, we propose to use timestamp and the location as pivots in generating contextual information from these two main sources (step 3). Public information can be any kind of information from documentary to news and personal information can be any private information from personal schedules, mails to personal achieves.

Once we have gathered rich contextual information from the two main sources, we perform an information extraction by analyzing the contextual information to get potential keywords for our photos (step 4). Some information like what happened, who were the famous people, and organizations in that event can be proposed to the users as our candidate keywords. In addition, we try to find the relationships between those keywords in order to improve the quality of our metadata in semantic way. A photo is worth a thousand words. Smeulders et al. define this semantic gap as: “the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation” [4]. In this sense, it is a must that user have to verify and validate each of the candidate keywords. However, once they are verified, they can be reused later.

Finally, we convert those keywords, their relationship and the already in place EXIF metadata into photo metadata that can be annotated to the photo (step 5).

## 4 Conclusions

We propose a novel approach in generating contextual metadata for photos that requires much less effort from the users’ side. We also believe that our approach is a new paradigm that can help cope with the semantic gap in personal photo management software. We are now building a prototype of our system to evaluate its effectiveness and accuracy.

Additionally, we plan to enhance our system with content based retrieval and users’ behaviors analysis to enjoy fully enriched and adaptive metadata.

## References

- [1] <http://www.flickr.com>
- [2] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. “Automatic organization for digital photographs with geographic coordinates”. In *4th ACM/IEEE-CS Joint Conf. on Digital Libraries, 2004*.
- [3] <http://www.exif.org>
- [4] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. “Content-based image retrieval at the end of the early years”. In *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349-1380, 2000.