Running title: Magician’s Hat

**Magician’s Hat: Master Our Collections of Photos**

Maoxu Li

Department of Computer Science, The George Washington University

lim@gwmail.gwu.edu

**Abstract**

People own more and more digital photos, which lead to a challenge of personal photo management. On one hand, many network photo services, especially those cloud-based ones, encourage people upload and share their photos. On the other hand, it is getting difficult to keep track on all of the distributed photos. For safety and management reasons, apart from the shared photo storage, people usually tend to maintain large centralized photo storage, which is established either on a personal device or on a cloud based network service. This separation in utilization of personal storage and network storage has increased the cost of personal photo management.

This paper proposed a solution that integrates personal storage and network storage to provide an advanced personal photo management with lower cost and enriched capabilities. The photos are distributed stored on personal devices and network services, while a central service is deployed to collect the information of photos globally. The system provides a unified view of personal photo management. Based on centralized photo files information and detection algorithm of identical or similar photos, advanced search supports SQL like syntax and content-based search, advanced backup supports a global duplicates view, excessive duplication removal, and scheduled backup.

A set of command-line desktop applications are designed and implemented as a proof of concept. More details will be studied and realized in the future work, and obviously the idea of this work can be extended to similar applications.

**Keywords:** photo management; photo search; photo backup; similar photo

**Table of Contents**

1 Introduction 4

2 Functional Analysis 6

2.1 Unified view of photo management 6

2.2 Advanced search 6

2.3 Advance backup management 6

3 System Design 6

3.1 System Logics 7

3.2 Software structure 7

4 Implementation 8

4.1 Platform 8

4.2 Software modules 9

4.3 Key technologies 9

5 Discussion 10

5.1 Future work 10

5.2 Conclusion 10

6 References 11

# Introduction

With the lower price of digital camera and camera becomes a regular built-in device in cell phone, people are owning more and more digital photos, not only photographers, collectors, and professional organizations, but also the individuals who love to take and share photos. During the past years, social networks have made it easy to manage and share a large amount of digital photos. Flickr and Facebook provide two of the most popular services that facilitate the backup and sharing of personal digital photos.

Many people take photos with the camera in cell phone every day and everywhere, some times with digital camera on trip, backup the photos to PC or laptop with large storage, then upload some favorite photos to one or more social networks from computers, even from cell phones directly. This is the general way for many people to manage and share their photos. Finally, people scattered their digital photos on their personal computers, devices, and social networks. It is getting difficult for people to keep track on all of their photos (Figure 1).

Figure 1. Personal photo management

Due to the management issues, together with safety considerations, some people tend to establish large centralized storage on a personal storage, and always backing up all photos on this storage. All photos distributed on other personal devices or network services are only duplicates out of the central storage. Some other people, however, choose a network service, like Flickr or Facebook photo, to establish the centralized photo storage. The drawbacks of these two strategies are obvious. They both ignore the utilization of storage out of the central storage, hence finally increase the cost of personal photo management, especially when the amount of photos gets huge. In addition, the safety of central storage is still an issue.

A straightforward idea to solve this problem is to integrate the utilizations of personal storage and network storage. However, some other issues have to be considered:

* Determine which locations a photo may be in before finding it.
* Track and make sure each of the photos has at least one copy.
* Manage the migration of storage or network services.
* Remove excessive duplicates of the photos.

In light of above idea, this paper proposed a central management system for personal photo management. A central service is deployed to collect the information of personal photos globally. Personal photo management, including advanced search and advanced backup management will be done with the assistance of the central service. Personal photos will be stored in personal devices or network services at convenience. The system provides a unified view of the personal photos and then management (Figure 2).

Figure 2. Personal photo management with centralized service

The following sections of this paper describe requirements analysis and system design in details. A set of command-line desktop applications is implemented as a proof of concept.

# Functional Analysis

## Unified view of photo management

The major difficulty of personal photo management is the separated views of the photos and their management due to the distributed storage locations. This system tends to offer a unified view of personal photo management based on the centralized service. User may access the system via different personal computers or devices, even public computers, to get a same global view of all photos and complete management operations within this view.

## Advanced search

Finding target photos is a big issue especially on a large amount of photos. Based on the photos information in central database, the system offers some advanced search capabilities beyond the fundamental search based on file name. The system provides search based on file stat and metadata with SQL like syntax. The system also provides content-based photo search to find identical or similar photos.

## Advance backup management

For safety, the photos should have more than one copy stored in different physical devices or network services. The system needs to provide a global view of duplicates and have capability to remove excessive duplicates. Duplicates may be photos that 100% identical or different slightly, for example, more white space, degraded quality or dimensions. The exact identical photos can be detected with simple scheme, for example the timestamp and size of the file, or the hash value of the file data. But the detection of similar photos, even with slight differences, is still a big challenge. The system will synchronize the photo files between personal devices and network services according to users’ schedule. The central service will mediate the files transfers between network services. However, the file transfers between personal devices and between personal device and network service may consider peer-to-peer manner that will reduce the workload of central service.

# System Design

The first step of system design is to determine the fundamental architecture. One possibility is establishing a web-based application. A web-based service maintains the database and provides a unified user interface with web pages. This architecture is suitable for providing personal photo management as a public service. Another possibility is realizing the whole system as desktop applications. This architecture makes the system as a personal software tool. Beside, the system may also be developed as a Facebook application, which will take the advantage of its existing social platform.

This work initially designs the whole system as desktop applications.

## System Logics

The system works based on the central service and the information of photo files stored in personal devices or network services. A file synchronizer module is deployed on local file system of personal devices to monitor and parse photo files. The resultant photo file information is transferred to and stored in the central database, which is maintained by the central service module. The photo files on network services are also monitor and parse with the file synchronizer module deployed in a personal device or a network server. The central service achieves advanced photo search and detection of identical or similar photos based on the data in database. To realize backup, the files may exchange between personal devices and network services. However, the central service only collects the information of photos to support the management functions. It does not store photo files (Figure 3).

Figure 3. System logics

## Software structure

A file synchronizer software module is designed with an outer interface named Synchronizer, through which the module receive commands from other synchronizer or central service. The core function of file synchronizer is extracting photo information, which includes some levels, for example, the simple information such as the timestamp or size of the file, the meta data of the photo, and the complicated feature data of the content of the photo used to do similarity analysis, or thumbnails of the photo. Particular software objects will be loaded by synchronizer to complete the interface functions, such as local file system monitor, file access, or network service access (Figure 4 left part).

A core service module implements two outer interfaces, one named Profile to receive photo information from file synchronizer modules and the other named Filter through which the user interface module search photos. The core service module maintains connection with database. Photo information is stored into database and the functions, such as photo search and content matching, are done based on the data in database (Figure 4 right part).

Figure 4. Software structure

# Implementation

A set of command-line desktop applications is implemented as a proof of concept. The implementation focuses on the application framework and the key functions, including photo file parsing, search, and detection of identical and similar photos.

## Platform

The implementation is on below running and development platforms:

* Mac OS X 10.6
* XCode 3.2
* C++, STL
* SQLite3
* OTL
* ZeroC Ice,
* OpenCV

## Software modules

The implementation includes three command-line applications:

* hatcore: Central service with database
* hatsync: File synchronizer
* hat: User interface

## Key technologies

Current implementation only completed part functions of the system. Here list some key points of programming or image processing techniques.

* **Monitor file system changes**

All major operating systems provide APIs for file system monitor. On Mac OS X, there are two levels of file system event API: fsevents provides a system level mechanism of file system monitor with a special virtual device. It can capture file system events on file level. File system evens API is a high level wrapper to provide application level file system monitor. It capture file system events on directory level, hence the actual file changes need to be analyzed by application. This implementation uses file system evens API to monitor directory level changes, then checks the timestamp and size of the files to determine their changes.

* **Extract image metadata**

Regular image file formats, such as JPEG, contain extra data to describe the information related with the photo, such as the dimension, the original time it is taken, even GPS information where it is take. Most of the popular image processing libraries, such as ImageMagick, FreeImage, and OpenCV, provide such API to read or write the metadata in photo files.

* **Detect identical or similar image**

According to above design, finding identical or similar photos is a key function in the system. Identical photos can be detected with simple scheme, for example, checking the timestamp and size of the file, or the hash value of file data. However, detection of similar photos, even with slight differences, is a challenge task. The general approach is extracting some feature data of the photos then calculating the distance of the feature data. A mostly used simple algorithm is comparing the gray or color histograms of two photos. Many feature data of photo can be calculated with the functions in image processing library like openCV, the complicated and effective algorithm can be designed to detect similar photos faster and more precise.

* **Network services access**

The system integrates the utilization of personal storage and network storage. Current implementation considers Flickr and Facebook photo services. Both services provide their dedicate APIs, which are similar in general. They both work on REST protocol that transfer message via HTTP session. In theory, any programming language that supports IP connection can be used to access the services. Some wrapper implementations with certain language can be used to simplify the development.

# Discussion

## Future work

Current work gave a fundamental design and implementation of the system as a set of desktop applications. The system then works as a personal software tool. In the future work, a web-based application will be designed and implemented to provide personal photo management service.

Based on the study of advanced algorithm to detect similar photos, advanced search based on content matching will be discussed in the future work. Advanced backup management, including global duplicates view, excessive duplication removal and scheduled backup will also be discussed in the future work. The asymmetric file synchronization between personal devices or network services, including peer-to-peer file exchange may be considered also.

## Conclusion

Centralized personal photo storage can’t satisfy the growing requirements. This paper proposed a solution that integrates personal storage and network storage to provide an advanced personal photo management with lower cost and enriched capabilities. The photos are distributed stored on personal devices and network services, while a central service is deployed to collect the information of photos globally. The system provides a unified view of personal photo management. Based on centralized photo files information and detection algorithm of identical or similar photos, advanced search supports SQL like syntax and content-based search, advanced backup supports a global duplicates view, excessive duplication removal, and scheduled backup. A set of command-line desktop applications are designed and implemented as a proof of concept. More details will be studied and realized in the future work, and obviously the idea of this work can be extended to similar applications.

# References

1. <http://www.flickr.com>
2. <http://www.facebook.com>
3. <http://www.zeroc.com>
4. <http://www.sqlite.org>
5. <http://otl.sourceforge.net>
6. <http://opencv.willowgarage.com/wiki/>