Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

A Framework for Attention-Based Personal Photo Manager

Wen-Hung Liao
Department of Computer Science
National Chengchi University
Taipei, TAIWAN
whliao@cs.nccu.edu.tw

***Abstract*—In this paper, we propose a novel framework for the design and implementation of an attention-based personal digital photo browsing platform. The key concept that separates the proposed system from existing ones is the incorporation of user interaction patterns to infer the level of interest in a particular photo. Specifically, we use web cameras to record and analyze the viewing behavior of the user and attempt to correlate the interest of the viewer to the effective viewing time. We also devise an updating scheme to efficiently renew the timing parameter. To build a comprehensive photo browser, external EXIF data and face detection results are utilized to coarsely classify the digital images. Moreover, measures of image quality, including sharpness and contrast, are calculated to rank the search results. Finally, a ranking-based algorithm is utilized to integrate the clues acquired from different modules.**

***Keywords*—attention-based photo manager, image quality analysis, user behavior.**

I. Introduction

The proliferation of digital cameras in recent years has generated serious data management issues due to the quick accumulation of large amount of digital photos. Compared to conventional film-based photos, digital images are inexpensive to acquire and store. It is therefore common to have thousands of personal photos on a storage device, either locally on a PC or remotely on a web site. These photos are often unorganized because of lack of convenient photo management software. They are, at best, arranged by folder or time of creation. When the number of photos increased over time, efficient indexing and searching become difficult unless proper annotation is attached to the image beforehand.

Currently, there are two main types of digital photo management software in widespread use. PC-based solutions usually come with viewing and basic editing capabilities. They also provide directory-based structure and thumbnail images for quick browsing. Web-based albums, on the other hand, focus on sharing. The task of organizing images into different groups is left to the user. Both approaches offer users the option to add keywords or comments to individual photos, although this function is seldom utilized in practice due to the effort involved in manual annotation.

To achieve automatic organization of digital photos, external information such as file name, size, date of creation/modification, and EXIF can be employed. EXIF tags are especially useful since they provide valuable information regarding the imaging formation process, including shutter speed, flash condition, and focal length. Content-based image retrieval systems that rely on internal information are also gaining popularity [1]. Attributes such as dominant colors, color histogram and distribution are some commonly used features in query processing. High level image understanding has the potential to enhance the performance of the photo management system. However, most systems utilize simple primitives since robust segmentation and interpretation of scene content is still a tough task.

In this paper, we are targeting at automatic management of personal photos. Personal photo collection is different from image database in several ways. Firstly, average users normally don't spend a lot of time organizing their digital photos. It is rather demanding to ask the user to add remarks to their photos for indexing and future search purposes. Therefore, in the proposed photo manager, we will not rely on keyword-based search. Secondly, images of personal acquisition are not guaranteed to possess good quality. Mass storage device is nowadays so inexpensive that most users just copy all the files on the storage card to the hard disk without filtering. In this process, photos of poor quality are usually left intact. An effective photo manager must be able to evaluate image quality in an objective manner so that higher quality images will possess higher ranks. Finally, personal photos can generally be classified to several basic types according to the purpose of shooting. Coarse classification is feasible by applying object recognition techniques. Image databases, on the other hand, can contain pictures of very distinct nature, making automatic classification much more difficult.

An important aspect in the design of personal photo browser that is until now neglected is user behavior. Attention level can be an effective indicator of the observer's interest in a certain picture. Longer viewing time usually signals stronger interest. This human-computer interaction process, when properly recorded and interpreted, can play an important role in organizing and personalizing the photo collection. The proposed framework will attempt to exploit this feature to boost system performance.

It is argued that digital photos are shared in multiple forms and viewed in many different ways at present days. Therefore it

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

may not always be feasible to record the interaction process and associate it with a photo collection. However, most people do keep a local collection of 'core' images which are either taken by the owner or received from close friends. This is the kind of photo album that suits the design introduced in this paper.

To summarize, the main objective of the research described in this paper is to propose the framework of a flexible and intelligent photo management system that can:

- gather image information for user reference automatically
- coarsely classify images according to image content
- evaluate importance of each image by image quality and user interaction patterns
- perform integration of multiple cues for preference ranking.

The rest of this paper is organized follows. In Section 2 we present the framework of the proposed photo management system and outline the key components. Section 3 discusses the integration of evidences from different sources using a ranking strategy. Conclusions and future improvements are briefly summarized in Section 4.

II. THE PROPOSED FRAMEWORK

This section describes the overall framework of the proposed personal photo management system. It contains three major components, namely, 1) information extraction module, 2) image quality evaluation module and 3) user behavior analysis module, as illustrated in Fig. 1. The information extraction module has two sub-systems: one for the collection of external information such as EXIF, the other for obtaining internal description using simple classification techniques. This part of the system forms the basis for query. It is similar to most database systems, except that the proposed method is enhanced with image content analysis results. The image quality evaluation module attempts to estimate an index that is directly linked to the quality of the digital photo based on measurements such as contrast and sharpness. The user attention modeling module sets to measure user interests by analyzing viewing behavior. Effective viewing time can be estimated accurately if a web camera is available. It can also be approximated by analyzing user interaction patterns.The main function of these two modules is to rank the search results. They can work independently or collaboratively. An information fusion system is utilized to integrate the results when necessary. We give detailed descriptions of the constituent modules in the following.

A. *Information Extraction Module*

As stated previously, the information extraction module is responsible for providing data that will be used for query purposes. In the proposed system, both external and internal information will be exploited. Traditionally, external information is mainly concerned with file-level structure such as file name, size, date and type. If the photo is taken by a digital camera, however, EXIF tags will be available. EXIF, a standard format in digital photo which is embedded in the JPEG header, records the condition of image formation. This information will help to identify important parameters such as exposure time, flashing mode and original creation time. The desired data can be conveniently extracted by parsing the EXIF. Newer cameras with GPS function also record the coordinates in EXIF. The location can be easily identified in a similar manner.

Internal information is derived from image content and structure. It is usually more difficult and time-consuming to compute than external information, especially when the image collection contains photos taken in different scenarios. Automatic photo tagging has the potential to provide cues for indexing [2]. But this technique is still evolving and improving. In our current implementation, we rely on a more robust approach: face detection.

The face detector we developed in [3] can identify faces and their locations in a photo with high accuracy. With this information, the photo can be roughly classified into four groups: (1) portrait: close-up shot of a single person, (2) group photo: pictures that contain more than one person. Photos in this category can be further classified according to the number of people inside, (3) tourist photo: photos that contain one or two persons, with emphasis on the background objects, (4) scenery photo: photos that contain no detectable face. It should be noted that the above classification is by no means complete, but should serve the purpose of illustrating the key concepts in the proposed architecture. Advances in object class recognition will undoubtedly increase the number of identifiable categories in scene classification results [4].

B. *Image Quality Assessment Module*

It is frequent to have images of poor quality in a personal photo collection as a result of wrong exposure, inaccurate focus or motion blur. These types of pictures are usually excluded when building an image database. However, average users generally do not spend the effort to dispose of these seemingly harmless pictures, as the storage cost has decreased significantly. It is up to the proposed photo manager to formulate some criteria to sort them out.

The search for a universal index to measure image quality has been futile to date, especially in the no-reference case [5]. There have been some recent progress in image quality assessment [6], yet the universality of the proposed measure requires further validation. Instead of finding a single parameter to characterize the quality of a given picture, we perform the evaluation from two aspects: sharpness measure and exposure condition.

We propose to compute the sharpness metric to from edge distribution. The basic idea is that well-focused images usually contain clearly defined edges. Therefore, if we apply edge detector and retain the same amount of edges, we will get more strong and isolated peaks in the edge map for sharp images than those of the blurred images. Fig. 2 (a) and (b) show the Sobel edge map of an original and Gaussian-blurred image, respectively. It can be seen that the edges of a blurred image tend to cluster. As a result, if we retain a fixed percentage of the edge pixel, say q, and calculate the *effective number of* *neighbors (ENN)* for each edge point *p* with in a neighborhood N_p (usually a *dxd* window) according to:

$$ENN(p) = \sum_{p' \in N_p, p' \neq p} \frac{I(p')}{d(p, p')} \tag{1}$$

where $d(p, p') = |x - x'| + |y - y'|$ (2)

and $I(p') = \begin{cases} 1, & \text{if } p' \text{ is an edge pixel} \\ 0, & \text{otherwise} \end{cases}$ (3)

then the out-of-focus images will generally produce larger values of ENN because of the clustering phenomenon. Notice that denominator of Eq. (1) is the Manhattan distance so that farther neighbors will get less weight, hence the term: effective number of neighbors. It has the additional advantage of reducing the dependency of ENN on a particular choice of *d* (window size defining the neighborhood). Using *q=5%* and a neighborhood of *15x15*, we obtain the ENN for Fig. 2(a):7.35, and Fig. 2(b):16, respectively.

Table I summarizes the average ENN and its standard deviation of two sets of photos (focused vs. out-of-focus), each containing 100 digital images. A preliminary conclusion is that the metric (ENN) performs rather well when the blurring is global.

TABLE I. ENN OF FOCUSED AND BLURRED IMAGE SET

Image Type	Average ENN	Standard Deviation
Focused	9.2	2.9
Out-of-focus	14.9	1.3

Exposure condition can influence the quality of the captured image in a significant manner. Wrong exposure settings usually result in photos whose histograms are either concentrated in the lower (under-exposure) or higher (over-exposure) ranges. But it is also difficult to find a universal index to determine whether a photo has proper lighting due to the dependency on image content. In the current implementation, we utilize the entropy of the 'value' histogram in the HSV color system to perform the evaluation. Photos with lighting problems generally have lower entropy than well-lighted ones. For images using 8-bit representation, the entropy is between 0 and 8. The result can be normalized to [0,1] in a straightforward manner.

Both sharpness measure and exposure condition are global metrics, meaning that the computation is carried out over the whole image. Sometimes the defect is local. For example, under-exposed subject in a backlighting condition or red-eyes using flash. We are aware of these issues, and will incorporate these local, content-dependent features in the future work.

C. User Behavior Module

A critical component that separates our system from other photo browser is the user behavior module. The proposed system takes into account user interaction pattern and applies this information to label viewer's interest in a particular photo. When a query is made, photos that have received more prior attention should be presented first.

In the simplest case, we can record the viewing frequency and time (T_v) during browsing to indicate user interest. However, the system can not tell if the user is actually viewing, especially in the slideshow mode. A more sophisticated approach would require a web camera. Using the same face tracker discussed previously, we can detect the presence of the user and his/her attention level. In the current version, the attention level is set to be proportional to the size of the detected face. Combining these data, we can derive the 'effective viewing time' (*T*) as illustrated Fig. 3 (area under the curve). We set an upper bound on T, considering that viewing for 40 seconds is the same as viewing for say, 50 seconds. Hence, the effective viewing time can be normalized to the range [0,1]. When a web camera is not available, we will simply set the effective viewing time according to Eq. (4):

$$T = k \times T_v \quad \text{where } 0 < \text{k} < 1 \tag{4}$$

Although the effective viewing time better reflects the user's interaction with the photo, it favors old pictures as them have more chances to be looked at. A new entry will have zero effective viewing time when it is first introduced to the collection. To address this issue, we will designate a decay factor for each photo. The underlying assumption is that a newly entered instance will receive a high score to signify its importance. On the other hand, an image's significance will decrease if it has not been viewed as much compared to others. Denote the significance measure as S ($0 < S \leq 1$),), we will update its value according to Eq. (5):

$$S^i(t+1) = S^i(t) \bullet \alpha^{\beta(T_i - \bar{T})} \tag{5}$$

where, Ti stands for the effective viewing time for the ith image, $\bar{T}$ is the average effective viewing time, α and β are parameters to adjust the rate of decay (α is set to 1.2, β is set to 2 in our prototype system.) $S^i(0)$ is initialized to 1. It is set to 1 if it Eq. (5) returns a value greater than 1. As the browsing progresses, *S* is re-evaluated. If the current photo has a shorter-than-average viewing time, its significance will decrease according to Eq. (5). Conversely, the significance increases if the photo receives more attention.

III. FUSION OF MULTIPLE CUES

The proposed framework suggests that information from different modules should be integrated to enhance performance and flexibility of query. The information fusion can be done manually, meaning that users are allowed to set the preferences by assigning different weights to the sharpness measure module, the exposure condition module and the user behavior module, respectively. Another approach is to learn the weights from user's ranking.

First, we present *n* photos to the user and ask him/her to rank them according to personal preference. We then re-order the photos based on user ranking. The features (output from sharpness module, exposure module and attention module in our case) associated with each photo form a row in the following matrix:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdot & \cdot & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdot & \cdot & x_{2,d} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n,1} & x_{n,2} & \cdot & \cdot & x_{n,d} \end{bmatrix}$$

The contribution from each feature is modeled by $w^T=[w_1, w_2,...,w_d]$. The weight assignment can be formulated as the solution of:

$$Xw = R \qquad (3)$$

where R is a n-dimensional vector whose entry value decreases as the index increases. For example, we can set $R^T=[n, n\text{-}1,...,1]$. A better practice, though, is to incorporate some non-linearity in R (e.g., $R^T=[n^2, (n\text{-}1)^2,...,1]$), so that the difference between the top-ranked photos will be amplified to a larger extent.

With the information fusion module in place, it is possible to perform some 'non-traditional' queries such as:

> Search for "good" (quality evaluation) photos which were taken during last July (time information from EXIF) and contain more than one person in them (content-based classification), rank the results automatically (ranking by fusion of attention level and image quality).

A screen shot illustrating the user interface of the prototype is presented in Fig. 4.

IV. CONCLUSIONS

In this paper, we have proposed a novel framework for personal photo browsing by integration of multiple-cues, including image content, image quality, and user behavior. We have proposed an effective measure to quantify image sharpness. We have also devised a method to correlate user behavior to a 'significance' index A rank-based mechanism has been developed to enable integration of multiple cues. Finally, a prototype has been built to assist initial evaluation. Further refinements to the processing modules, as well as extensive user testing will be conducted to guarantee satisfactory user experience

REFERENCES

[1] A.W.M. Smeulders, M.Worring, S.Santini, A.Gupta and R.Jain, "Content-Based Image Retrieval at the End of the Early Years", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, pp.1349-1380, 2000.

[2] Jia Li and James Z. Wang, "Real-time Computerized Annotation of Pictures," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 6, pp. 985-1002, 2008

[3] W. Liao, T. Wang and Y. Lin, "Robust Multi-pose Face Detection in Video", Proceedings of the 20th Conference on Computer Vision, Graphics and Image Processing, Aug. 2007.

[4] Bosch, A. , Zisserman, A. and Munoz, X., "Scene Classification Using a Hybrid Generative/Discriminative Approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, no.4, pp 712-727, 2008.

[5] Z. Wang, A. C. Bovik and L. Lu, "Why Is Image Quality Assessment So Difficult?", Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing, Vol: 4, pp:3313 -3316, 2002.

[6] Srenivas Varadarajan, Lina J. Karam, "An Improved Perception-based No-reference Objective Image Sharpness Metric Using Iterative Edge Refinement", Proceedings of the International Conference on Image Processing, pp. 401-404, 2008.

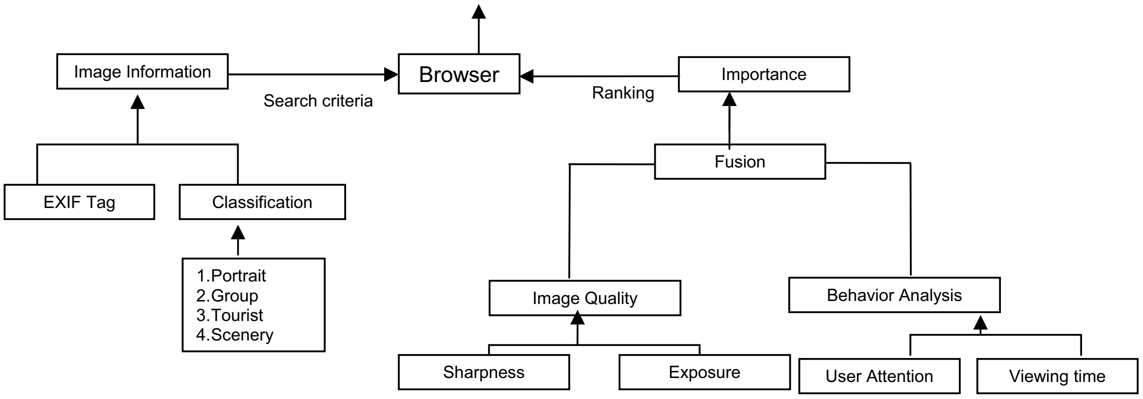

Figure 1. Overall framework of the proposed personal photo management system

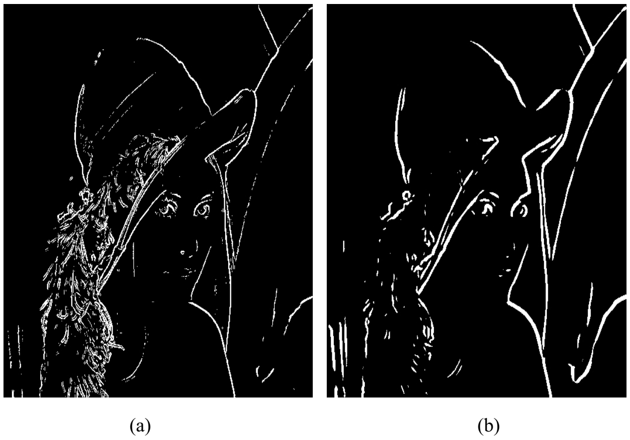

Figure 2. Sobel edge map of the original Lena image (a) and the blurred version (b).

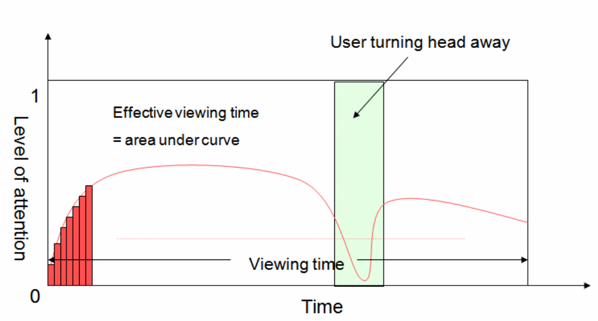

Figure 3. Computing the effective viewing time

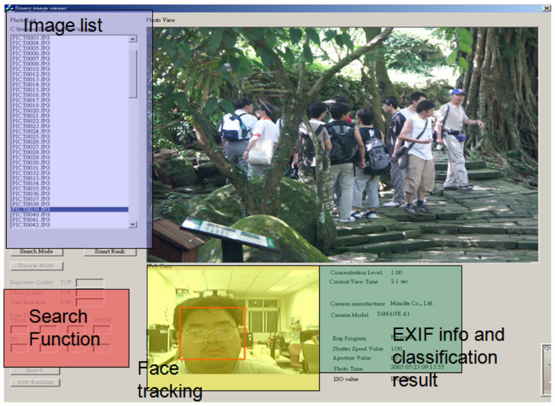

Figure 4. User interface of the proposed photo browser