Embedded Tags and Visual Querying for Face Photo Retrieval

Chaminda de Silva, Toshihiko Yamasaki, Kiyoharu Aizawa

Department of Information and Communication Engineering
The University of Tokyo, 102B2, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{chamds,yamasaki,aizawa}@hal.t.u-tokyo.ac.jp

Abstract. We investigate the utility of automated digital photo retrieval using *faceboxes*. The sizes and locations of faces in each photo are automatically detected and embedded within the image file. An interactive user interface allows querying for photos visually, in a simple and intuitive manner. We present a user study conducted for evaluating the system on a personal collection of 10,000 digital photos, and report the results. The average search time, including time for sketching and browsing the results, is 31.2 seconds. Usability ratings indicate that the interface is easy to use, and useful as a tool for photo retrieval.

1 Introduction

There has been a high growth in the amount of digital photos acquired and stored during the past few years. However, progress on systems for effective organization of digital photo collections and automated photo retrieval has been relatively low. While a number of free and commercial photo manager programs exist, most digital camera users simply copy their photos in to a hierarchy of directories. Most photo menagement software require the users to annotate images using keywords (commonly known as 'tags') for faster retrieval. However, users find it tedious to annotate their photos [8, 4]. Further, the tags are stored centrally with the software and will not be propagated when the photo is uploaded or sent to a different location.

Because of these difficulties, Content Based Image Retrieval(CBIR) has become a growing research topic. The common approach in CBIR is to analyze the images first and generate textual metadata regarding the content, for search by users. However, since it is difficult to specify an image with a textual description, visual querying techniques seem more prospective. While querying by skecthes seems a good approach, its effectiveness is limited due to the time consumed and sensitivity to color differences [10].

Faces are an important category of content in photographs. This fact is utilized by several systems, both hardware and software. Digital cameras detect faces in what is seen through the lens, and adjust camera settings to ensure that the faces are in focus and well exposed. Social networks facilitate marking "faceboxes" in images for associating photos with people. We believe that systems for image retrieval will benefit from similar emphasis on faces.

In this paper, we propose a system that utilizes faces as a cue for content based retrieval of digital photographic images. A collection of information about faces contained in an image forms the metadata. These data are embedded in the image itself, in a compact and unambiguous format. Embedding metadata makes images "ready-to-search", eliminating the need of regenerating metadata when passed to a different computer. A user can retrieve such images by composing an iconic sketch with simple inputs, rather than sketching in detail. The system extracts the required search parameters from the sketch and searches the metadata, and retrieves similar results. We conduct a user study to evaluate the system and obtain feedback and suggestions for improvement.

The rest of the paper is organized as follows: Section 2 is a brief review of related work; Section 3 describes the functional components of the system in detail; Section 4 presents the user study and the results; Section 5 contains a brief discussion regarding the design and the results of the user study; Section 6 concludes the paper with suggestions for future directions.

2 Related Work

A detailed, recent review of the state of the art of content-based image retrieval can be found in [4]. Despite the large amount of research, real-world application of the technology resulting from such research is currently limited [4]. There has been an increase in research on user interfaces for CBIR. Bird et al. [1] propose a set of design considerations to facilitate content-based queries. Hove [7] studied how users translate information needs to visual queries, and showed that users tend to prefer iconic sketches to free hand sketches when the complexity of an image retrieval taks increases. Rouw, in his *PhotoIndex* system, uses iconic composition of visual queries for CBIR. Chang et al. [2] propose *Semantic Visual Templates* generated based on interactive queries for retrieving images from a large, unanotated image database.

There have been some efforts to retrieve images based on faces. FACERET [11] is an interactive face retrieval system based on matching faces using self-organizing maps. Girgensohn et al. [5] combine face detection and face recognition to achive semi-automatci labeling of images by person.

Recent user studies on how people store, manage and search for digital photographs show that most people organize their photos into a hierarchy of folders, where the folder names contain information about the time, location and event where the photo was captured [9][8][10]. Westman et al. [13] study and compare the image searching behavior of journalists and image archivers.

3 System Description

Figure 1 is an outline of the proposed system. The system consists of two subsystms. The first subsystem detects faces in each image, and embeds metadata in the image itself regarding the number and layout of detected faces. The second subsystem accepts user queries in terms of an iconic sketch, and queries

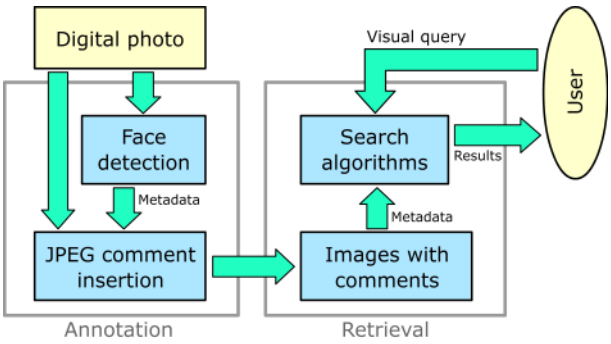

Fig. 1. System overview.

the images that match the parameters extracted from the sketch. The following subsections describe these subsystems and the algorithms in detail.

3.1 Face Detection and Image Annotation

For each image, the system performs face detection to identify the attributes of faces present in the images. For each face, a square "facebox" is detected, and its location and width(=height) is recorded. Figure 2 shows an image and the attributes of the detected faceboxes. The system uses an open source implementation of Viola-Jones feature detector[12] for face detection.

After face detection, images are stored as JPEG files, and the metadata are stored in the *comments* area of the JPEG file as ASCII text. The following values are stored, separated by spaces:

1. Preamble: '*FaceSrch*'
2. Width of the image in pixels
3. Height of the image in pixels
4. No. of faces (integer ≥ 0)
5. A sequence of attributes of faceboxes, separated by spaces, in the format

$$X_1\ Y_1\ W_1\ X_2\ Y_2\ W_2\ ...\ X_n\ Y_n\ W_n$$

where (X_i, Y_i) is the pixel coordinate of the top left corner of the i^{th} facebox, W_i is the width of the i^{th} facebox, and $i = \{1, 2, ..., n\}$.

Given that the maximum length of a JPEG comment is 65536 bytes [6], this encoding scheme can accomodate attributes of more than 1000 faces (sufficient for practical situations). For the image in Figure 2, the metadata will be:

FaceSrch 960 720 2 140 156 126 451 277 144

The interface to this subsystem allows the user to specify either a single image or a hierarchy of folders containing images, as input for annotation. It

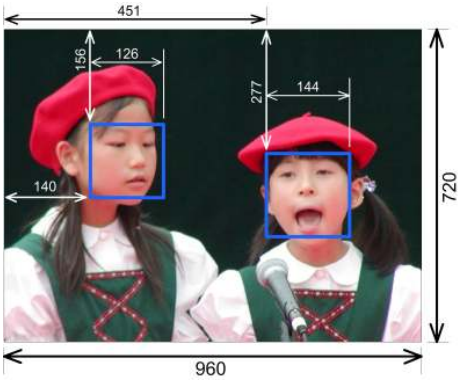

Fig. 2. An image with faceboxes.

is possible to automatically annotate images using a background process when new images are copied to the computer (for example, uploaded from a digital camera).

3.2 User Interaction Strategy

We employ a user interaction strategy that allows a user to form detailed queries quickly, with simple actions. The user starts by specifying a disk drive or a hierarchy of folders containing images to be searched. After selecting the folder/s, the Thumbnails tab allows the user to browse the collection of photos contained in them, as in many image viewing software.

The user can now start composing his query, to retrieve the desired images from this collection. The orientation of the image and the number of faces in the image can be selected using radio buttons and combo boxes (Figure 3b). The interface allows incomplete and low precision inputs, such as "any orientation", and "more than 3 faces". If the user remembers the layout of the desired image (with respect to the faces), or wishes to retrieve images with a particular layout, he/she can specify the image layout by making an iconic sketch. Upon selecting "Specify" option for face locations and sizes, a canvas initialized according to the selected orientation and the number of facesis shown to the user. The user can drag the faces inside the canvas to change their locations. The size of a face can be changed using two steps. A face is selected by placing the mouse cursor over it and clicking the left mouse button. If a mouse wheel is available, rotating it will change the size of the face. In case the mouse does not have a wheel, the slider below the canvas can be used for changing the size. If the user is more certain/particular about the locations of faces than their sizes, or vice versa, the "Search priority" slider can be adjusted according to the preference. After selecting the desired options and completing the sketch, the user clicks the "Search" button to retrieve images. The following section describes the algorithms used for searching the collection of images based on the user's inputs.

3.3 Search Algorithm

Upon selecting a folder, the system calculates the following attributes for each image using the information embedded in the the comment area as described in Section 3.1:

1. Height of the image H, in pixels
2. Width of the image W, in pixels
3. Number of faces contained in the image, n
4. For each face, the pixel coordinates of the center of the face box (x_i, y_i) where $i = 1, 2, ..., n$
5. For each face, the width of the face box

After the user clicks the "Search" button, the search algorithm is activated. First, the aspect ratio A of the image is calculated as $A = H/W$. Depending on the orientation the user selected, the appropriate set of images is selected as follows:

1. If the user selected "Portrait", select images where $A > 1$
2. If the user selected "Landscape", select images where $A < 1$
3. If the user selected "Not sure/Any", select all images

The next step is to filter the images according to the number of faces that the user has specified. This is straightforward since the number of face in each image is already known. For instance, if the user specifies "exactly 2" as the number of faces in he photograph, the system scans the list of images and selects only those with $n = 2$. If the user has not specified the layout of the photo, the results can now be displayed. If the layout has been sketched, a search algorithm is used to order the photos so that the images that are most similar to the logical sketch appear first within the search results. The "Search Priority" slider controls the weights of the algorithm to adjust the relative influence of face sizes and locations when estimating similarity. The results are ordered according to the descending order of similarity, grouped into sets of 12 thumbnails each, and shown to the user. The user can click on a thumbnail to get a larger view of the image in the "Image" tab, and also save it to another location if necessary.

3.4 Example Scenario of Retrieval

Figure 3 illustrates how the interface is used to retrieve a desired image(Figure 3a) from a collection of approximately 7000 photos stored in a folder hierarchy under "C:\FaceCoded". Figure 3b shows the user inputs for the query. Figure 3c shows that the image has been retrieved as the best match out of 247 portraits with two faces.

4 User Study and Results

We conducted a user study based on the proposed system with the following objectives:

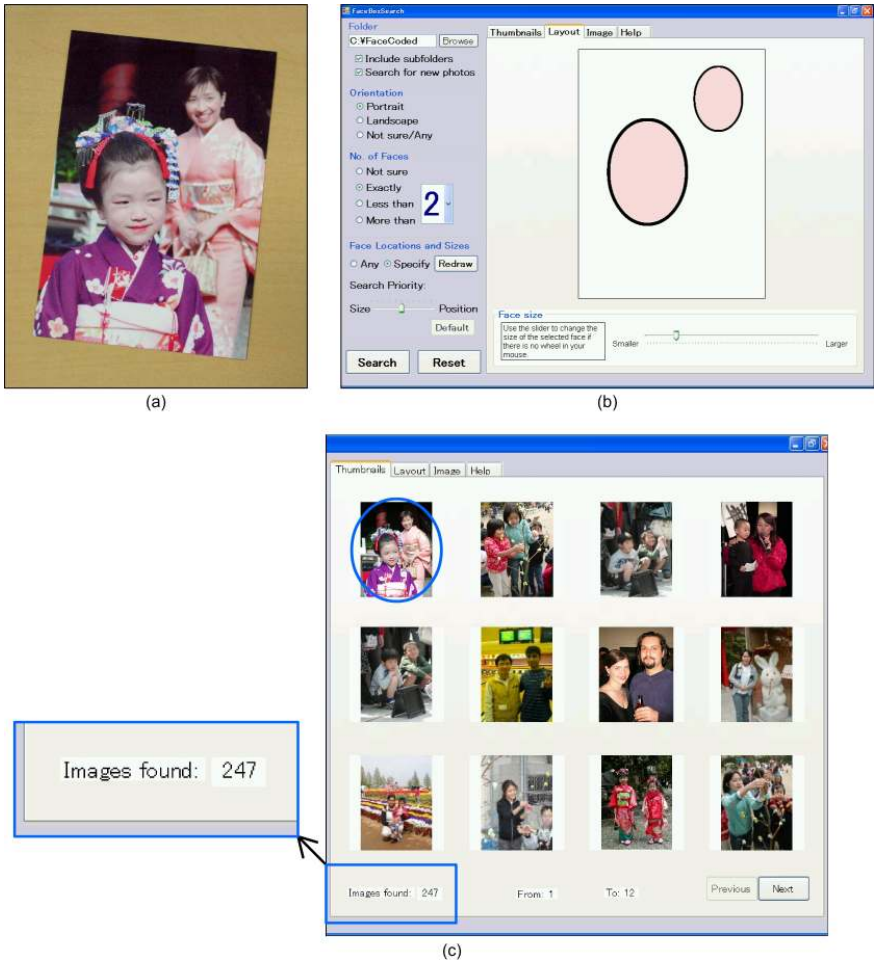

Fig. 3. Using the interface for image retrieval.

1. Identify the methods used by the ordinary digital camera users for organizing their photo collections
2. Evaluate the proposed system both quantitatively and qualitatively, in terms of search time and usability
3. Acquire feedback and comments in order to identify changes and future directions

The following sections describe this user study and the results.

4.1 Content and Procedure

The user study consisted of four sections. In the first section, the subjects answer a questionnaire regarding how they capture, organize and search for digital pho-

tographs. In the second section, the subjects use the proposed system to retrieve 12 *desired images* from a collection of 10134 digital photos taken by one of the authors. The desired images are grouped into two *sets* of six images each. One set is retrieved by specifying the layout while looking at a printed copy of the photo. The other set is retrieved by entering the layout as the user remembers. We followed the guidelines specified in [9] to simulate the task of searching for a familiar image, by showing the images in this set before the task but only for 10 seconds. The third section consisted of a usability study of the proposed user interface and user interaction strategy, based on the evaluation questionnaire proposed by Chin et al. [3]. In the fourth section, the subjects provided their feedback about the system and suggest improvements.

Sixteen voluntary subjects participated in the experiment. All subjects used some form of a digital camera, ranging from camera phones to Digital SLRs. All subjects were regular computer users. Two of the subjects were professional photographers. The two sets of desired images were alternated between subjects so that each desired image is retrieved using both methods during the experiment. The subjects took 25 to 45 minutes to complete the study, with an average time of 35.3 minutes. This time included short breaks between sections. The inputs and search times for retrieving images were recorded.

4.2 Results

According to the responses to the first section of the study, the subjects maintained digital photo collections with sizes ranging from 100 to 30,000 photos . Fourteen (83%) of the subjects had a collection of more than 100 photos, and eight (50%) had more than 1000 photos. Only two of the subjects used photo management software. All the subjects used a folder hierarchy for managing their photos, where folder names corresponded to details such as the date of capture, date of upload, the event, etc. The subjects who did not use a photo management software searched for photos by finding the relevant folder and then browsing the thumbnails. Only one subject named individual photos to facilitate faster search. 13 subjects said that they do not edit their photos at all. The results of the study confirmed observations by other researchers who conducted similar experiments [9][10].

For all the subjects and the images, the system was able to retrieve the desired image within the first 72 thumbnails (6 pages of results). The image was retrieved within the first 12 thumbnails in approximately 42% of the searches. The search time for a photo using visual queries ranged from 9.1 to 58.8 seconds, depending on the user and the complexity of the image layout. The average search time for retrieving an image at hand (set 1) was 30.1 seconds. The average search time for retrieving a familiar image (set 2) was 32.3 seconds. Related studies report an average time interval of 5.1 seconds to find an image from 80 thumbnails in one folder by browsing them [9]. Given that the proposed system searches 10,000 images in 98 folders, it is evident that the system makes image retrieval much faster. The system is especially useful for retrieval where the user does not know or remember sufficient information to narrow down the search into a few folders.

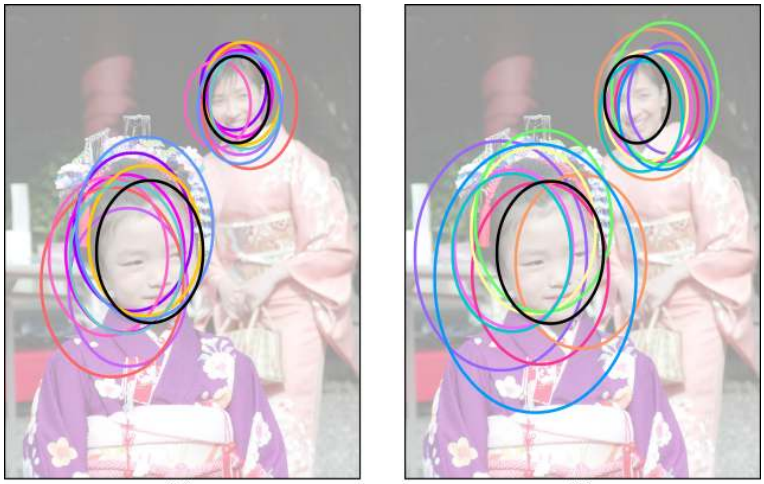

Fig. 4. Using the interface for image retrieval.

The difference between the search times for the two sets of images is quite small. This is mainly because the users took less time for sketching the query when the printed photo was not with them, and the system was able to retrieve the photos accurately even with imprecise sketches. Figure 4 compares iconic sketches made by the subjects to retrieve one of the test images. The ground truth for face locations (black) and iconic sketches made by subjects (other colors) are superimposed on images. Figure 4a contains iconic sketches made while looking at the desired image. Figure 4b contains iconic sketches made by what the subject remembered about the image. It is evident that the relative locations and relative sizes of faces were generally well remembered for a familiar image.

Below is a list of the criterion descriptor, response mean, mode (in parantheses) and the range of responses for the main criteria evaluated during the usability study. A seven-point response scale was used with 1 being the worst rating(very poor performance) and 7 the best (very good performance).

- Ease of submitting inputs: 6.7 (6) 5-7
- Organization of results : 6.1 (6) 5-7
- Ease of learning to use : 6.8 (7) 6-7
- Speed of retrieval : 6.7 (6) 6-7
- Reliability : 6.7 (6) 6-7
- Ease of use : 6.2 (6,7) 4-7
- Usefulness : 5.75 (5) 5-7

Asked what they liked about the software, most of the subjects stated that it provides an easy method to search for images. Some others stated that it made image search 'enjoyable'. Asked what they disliked about it, several sujects said

that they were not sure whether or how to use the "Search priority" slider. Some others stated that showing only 12 thumbnails is insufficient.

The subjects provided detailed comments and feedback on how to improve the system. While they found it easy to learn and use, they also desired additional functionality incorporated to the system. The most commonly suggested additional capabilities were face recognition, and face classification by gender or age group.

5 Discussion

About 10% of the photos in the collection used for the study included inaccurate face detections. However, only one user mentioned this as a bad point about the system. The reason for this might be the fact that people are used to browsing a large amount of photos to find a desired photo.

Three of the subjects said search using the software is enjoyable, while two others commented that they would like to have the software in their computers. This is encouraging, since a pleasant user experience is important in designing good user interfaces and interaction strategies.

According to the iconic sketches recorded during the user study, most of the users accurately remembered the face sizes and locations in relation to each other. The current search algorithm, which uses face sizes and locations in relaton to the canvas size, can be revised to make use of this and improve retrieval.

The prototype that we have developed uses two methods for face detection. When used in single image mode, the user can draw face boxes by clicking and dragging the mouse cursor on the image. Although manual annotation is not practical for a large number of images, this increases flexibility of the system and allows the user to annotate images where faces are difficult to be detected .

It is evident that the proposed user interaction strategy is more useful as a part of a more functional system for CBIR. For example, face recognition combined with the proposed scheme can be used both as a personal photo management system and a person-based image retrieval system. The system is particularly useful for graphic designers, *searchers* and *surfers* [4] who have a visual image of what they are searching for.

6 Conclusion and Future Work

We presented an encoding scheme for embedding metadata regarding human faces in JPEG photographs, and an interactive system for retrieval of photos using such metadata. The encoding scheme is sufficiently simple to be included in a digital camera capable of face detection. Embedded comments remove the burden of re-indexing images upon transfer, and computationally intensive image analysis during retrieval. The average search time for an image is 31.2 seconds. The user interaction strategy is intuitive, and easy to learn.

The current iconic sketches are designed for mouse. The interaction strategy can be revised to make it easier to use on pen-based devices such as tablet PCs.

We are woking on incorporating face recognition to the system, so that it can facilitate person-based photo retrieval.

Acknowledgments

We thank the voluntary subjects who took part in the experiment. Chaminda de Silva wishes to thank photographers Hiroshi Iguchi and Keiko Iguchi for helpful comments and additional feedback.

References

1. C. L. Bird, P. J. Elliott, E. Griffiths: User interfaces for content-based image retrieval. Intelligent Image Databases, IEE Colloquium on , vol., no., pp.8/1-8/4, 22 May 1996.
2. S. Chang, W. Chen, H. Sundaram: Semantic visual templates: Linking visual features to semantics. In Proceedings of IEEE International Conference on Image Processing 1998.
3. J. P. Chin, V. A. Diehl, K. L. Norman: Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. Proceedings of ACM CHI'88 Conference on Human Factors in Computing Systems, 1988 p.213-218.
4. R. Datta, D. Joshi, J. Li and J. Z. Wang: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys 40(2), April 2008. 5:1–5:59
5. Girgensohn, A., Adcock, J., and Wilcox, L. 2004. Leveraging face recognition technology to find and organize photos. In Proceedings of the 6th ACM SIGMM international Workshop on Multimedia information Retrieval (New York, NY, USA, October 15 - 16, 2004). MIR '04. ACM, New York, NY, 99-106.
6. E. Hamilton: JPEG File Interchange Format. C-cube Microsystems, Milpitas, CA 95035, 1992.
7. L-J. Hove: Evaluating Use of Interfaces for Visual Query Specification in Proceedings of NOBOKIT 2007, November 2007.
8. K. Rodden and K. Wood: How do People Manage Their Digital Photographs? in Proc. ACM Conference on Human Factors in Computing Systems (ACM CHI 2003), April 2003.
9. K. Rodden: Evaluating similarity Based Visualizations as Interfaces for Image Browsing. Technical Report UCAM-CL-TR-543, University of Cambridge, September 2002.
10. M. Rouw: Meaningful Image Spaces and Project PhotoIndex. Masters Thesis, Utrecht School of the Arts, Hilversum, the Netherlands, 2005.
11. J. Ruiz-del-Solar and P. Navarrete: FACERET: An Interactive Face Retrieval System Based on Self-Organizing Maps. In Proceedings of the international Conference on Image and Video Retrieval (July 18 - 19, 2002). Lecture Notes In Computer Science, vol. 2383. Springer-Verlag, London, 157-164.
12. P. Viola and M. Jones: Rapid object detection using a boosted cascade of simple features. In IEEE Conference on Computer Vision and Pattern Recognition CVPR), 2001.
13. S. Westman, P. Oittinen: Image retrieval by end-users and intermediaries in a journalistic work context. In Proceedings of the 1st International Conference on Information Interaction in Context, October 18 - 20, 2006). IIiX, vol. 176. 102-110.