Face Detection and Visual Querying for Interactive Photo Retrieval

Chaminda de Silva   Toshihiko Yamasaki   Kiyoharu Aizawa
Department of Information and Communication Engineering
The University of Tokyo
(chamds, yamasaki, aizawa)@hal.t.u-tokyo.ac.jp

ABSTRACT
We present a scheme for automated retrieval of digital photographs using *faceboxes*. The sizes and locations of faces in each photo are automatically detected and stored as metadata. An interactive user interface allows querying for photos visually, in a simple and intuitive manner. Two applications of the proposed scheme are demonstrated.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.1 [**Information Interfaces and Presentation**]: User Interfaces

Keywords
Facebox, multimedia interfaces, image retrieval

1. INTRODUCTION
There has been a rapid growth in the number of digital photos acquired and stored during the past few years. However, most people neither organize the photos they take (other than for copying into directories by date or event) nor 'tag' their photos [1, 3]. Therefore, accurate Content Based Image Retrieval (CBIR) will be very helpful for searching for photos. The common approach in CBIR is to analyze images and generate textual metadata for search by users. Visual querying seems more prospective due to the difficulty in specifying an image with a textual description. However, querying by detailed skecthes has limited effectiveness due to the time consumed and sensitivity to color differences [4].

"Face" is an important category of content in photos. Digital cameras utilize face detection to adjust camera settings to ensure that faces are in focus and well exposed. Social networks facilitate labeling faces in images, to associate photos with people. *Google Search* and *Picasa* allow filtering image search results by the presence of faces. We believe that systems for content based image retrieval will benefit from further emphasis on using faces as a search cue.
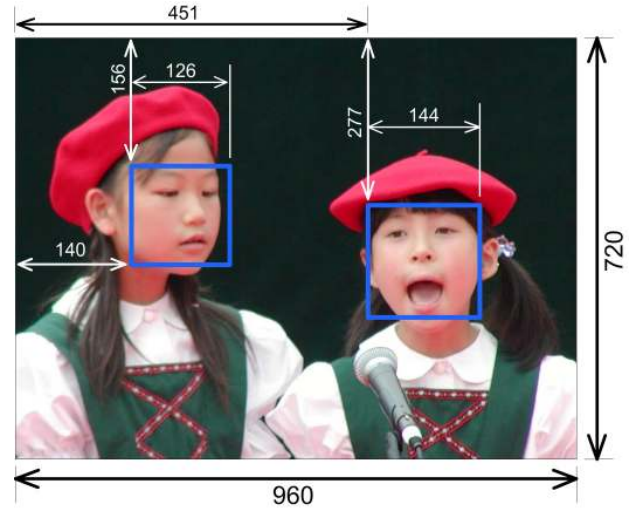

Figure 1: An image with faceboxes.

This paper presents an interactive system that utilizes faces as a cue for content based retrieval of digital photographs in JPEG format. It allows the user to describe a photo using a visual query that specifies the layout of faces in the photo, for more effective querying and retrieval. Such a system is particularly valuable for graphic designers, *searchers* and *surfers* [1] who have a visual image of what they are searching for. The proposed scheme consists of two subsystems. The first detects faces in an image, extracts their attributes and forms metadata for retrieval. The second provides a user interaction strategy that allows forming visual queries for image retrieval, and contains search algorithms for retrieving images that match the queries.

2. SYSTEM DESCRIPTION
2.1 Extracting Metadata
For each image, the system automaticaly detects faces in images using an open source implementation of *Viola-Jones* detector. The location and size of a square *facebox* bounding each face is determined. Figure 1 shows an image and the attributes of the detected faceboxes. These attributes are encoded into a tag and stored in the comments area of the JPEG image file. It is also possible to store the data externally as an index to a collection of images. The encoding scheme is sufficiently simple to be included in a digital camera capable of face detection.

2.2 User Interaction Strategy

We employ a user interaction strategy that allows fast forming of detailed queries. First, the orientation of the image, and the number of faces in the image are selected. Incomplete and low-precision inputs such as "any orientation", and "more than three faces" are permitted. If the user remembers the layout of the desired image (with respect to faces), or wishes to retrieve images with a particular layout, he/she can sketch an icon-based visual query of the image layout. If the user opts to specify the layout, a canvas with a set of face icons - initialized according to the selected orientation and the number of faces - appears on the screen. A face icon is selected by clicking the left mouse button on it. The user can drag an icon within the canvas to change its location. The size of a face can be changed either by rotating the mousewheel (if available) or by adjusting a slider. If the user is more certain/particular about the locations of the faces than their sizes, or vice versa, he/she can specify the search priority using a slider control in the user interface.

Upon submission of the search parameters, the system extracts the dimensions and metadata of the images in the search domain. These images are filtered by the orientation and the number of faces, if specified. The remaining candidate images are ordered according to their similarity to the face layout specified in the icon-based sketch, using a matching algorithm. The "Search priority" input 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 similarity, grouped into sets of thumbnails, and displayed to the user.

3. APPLICATIONS

The proposed interaction strategy is particularly useful for image search in professional work contexts such as journalism, where 40% of queries for images are related to persons, and the composition and layout are very important selection criteria [5]. We demonstrate two applications of the proposed scheme. The first is a prototype search engine that retrieves photos of famous persons. The search index contains metadata extracted using the method described in Section 1. The user can enter the name of the person together with a visual query, and retrieve images of the person with the specified layout. Figure 2a shows the inputs of an example query, and Figure 2b shows the best twelve matches out of 946 search results.

The second application is a personal photo organizer. Face detection is carried out as a background process when photos are copied to the computer from a digital camera. The metadata are stored in the *comments* area of the JPEG image files. Given that the maximum length of a JPEG comment is 65536 bytes [2], it is easy to accomodate attributes of more than 1000 faces (sufficient for practical situations). Having metadata within the image files themselves make them "ready to search" even when transferred to another computer. We demontrate this system on a collection of 10,000 photos from a personal collection.

4. CONCLUSION

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.
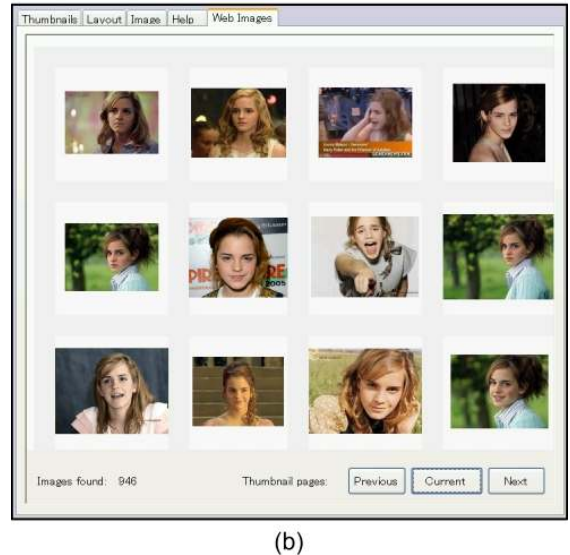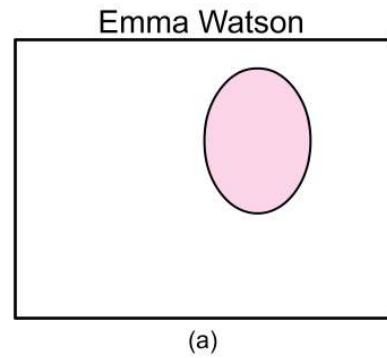

Figure 2: Using the interface for image retrieval.

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 intensive image analysis during retrieval. The user interaction strategy is intuitive, and easy to learn. We are working on improving the proposed scheme by incorporating face recognition.

5. REFERENCES

[1] 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):5:1–5:59, April 2008.

[2] E. Hamilton. *JPEG File Interchange Format*. C-cube Microsystems, Milpitas, CA95035, 1992.

[3] K. Rodden and K. R. Wood. User interfaces for content-based image retrieval. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 409–416. ACM, 2003.

[4] M. Rouw. *Meaningful image-spaces and project photoindex, Masters thesis*. Utrecht schools of the arts, Hilversum, the Netherlands, 2005.

[5] S. Westman and 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*, volume 176, pages 102–110, October 2006.