Face Detection and Visual Querying
for Editorial Photo Retrieval

Gamhewage C. de Silva
Interdisciplinary Initiative for Information Studies
The University of Tokyo
Hewlett-Packard Labs Japan
chamds@hal.t.u-tokyo.ac.jp

Kiyoharu Aizawa
Department of Information and
Communication Engineering
The University of Tokyo
aizawa@hal.t.u-tokyo.ac.jp

ABSTRACT

We present a system for automated image indexing and retrieval using *faceboxes*. The sizes and locations of faces in each photo are automatically detected, and stored together with keywords, as metadata. An interactive user interface allows querying for photos visually, in a simple and intuitive manner. The average times for indexing and searching for an image are approximately 20 s and 31 s respectively.

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, editorial photo retrieval

1. INTRODUCTION

Editorial photo retrieval is a specific application of image retrieval. Commercial photo providers, such as *Getty Images* and *Corbis*, facilitate automated retrieval by manually indexing images according to the content, composition and other criteria. Recent studies on image retrieval for professional tasks show that approximately 42% of queries are for photos of persons, and composition is considered the most important selection criterion for such images [4, 3]. At the current state of existing systems, it is possible to specify composition using pre-defined terms such as "head shot", "waist-up", and "full length". However, this method allows only a limited set of choices, and results in low precison. Further, labeling is performed manually, and therefore subjective and inconsistent. A photo labeled as "head shot" by one person might be labeled as "waist up" by another.

This paper presents an interactive system that utilizes faces as a cue for composition based indexing and retrieval of digital photographs. It allows the user to describe a photo using a visual query that specifies the composition of the photo,
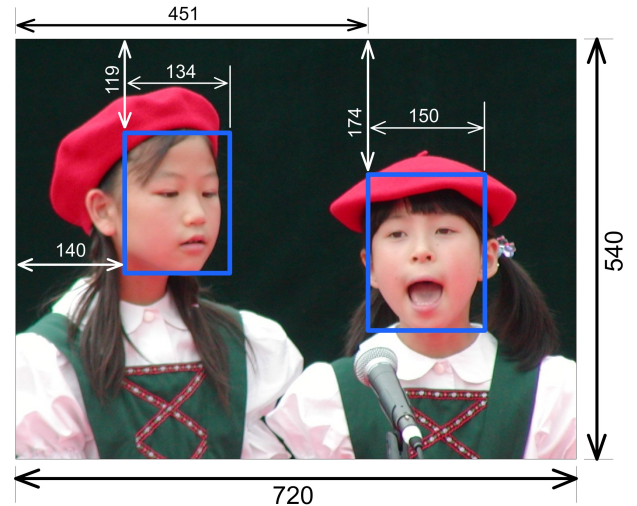

Figure 1: An image with faceboxes.

for more precise retrieval. The proposed system 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. IMAGE INDEXING

We propose an interactive, semi-automatic system for indexing images using facial composition. The user starts indexing by specifying an image or a directory containing images. The system automaticaly detects faces in each image. 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.

The user interface of the indexing system allows adding, removing and modifying faceboxes, if face detection has not been accurate. A facebox can be selected by left-clicking the mouse on it. The user can drag a facebox within the image to change its location. The size of a facebox can be changed by rotating the mousewheel (if available) or by adjusting a slider, after selecting it. A facebox is added by left-clicking the mouse on the image. Right-clicking on a facebox deletes it. The user can add names and keywords as additional metadata, to a single photo or all photos in the directory.

The attributes of faceboxes and the keywords are encoded to a textual tag in the following format:

FaceSrch <no. of faces> <location and width of each face-box> <keywords>

For example, the textual tag created for the photo shown in Figure 1 is:

FaceSrch 2 140,119,134 451,174,150 children,singing

The tag is stored in the *comments* area of the JPEG image file, so that it is not necessary to re-index images upon transferring to a different location (for example, downloading or sending by email). The metadata are also stored externally in a file, to be used as an index for faster searching.

3. IMAGE RETRIEVAL

The user starts querying for photos by entering person names and/or keywords. Thereafter, the orientation of the image, and the number of faces (persons) in the image are selected. Incomplete and low-precision inputs such as "any orientation", and "more than three faces" are permitted. The user can specify the desired composition using an iconic sketch [2]. 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 that match the keywords. These images are then filtered by orientation and number of faces, if specified. The remaining candidate images are ordered according to their similarity to the face layout specified in the iconic sketch, using a matching algorithm. The "Search priority" input controls the weights of this 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. Figure 2a shows the inputs of an example query. The query can be described in detail as "photos of Emma Watson, in landscape orientation, with the face to the right of the image.."etc. Figure 2b shows the best twelve matches out of 946 search results that match the given keyword, number of faces and orientation.

4. RESULTS

We demonstrate the system on a collection of 10,000 digital photos, from the personal collection of one of the authors and a free stock photography site [1]. Keywords were added to either batches of photos or individual photos, as appropriate. The average time for indexing an individual image, including the time for adding keywords, was approximately 20 seconds. Average search time for images containing 1-5 faces was approximately 31 seconds. An online version of the system is available at *http://www.faceboxsearch.com.*
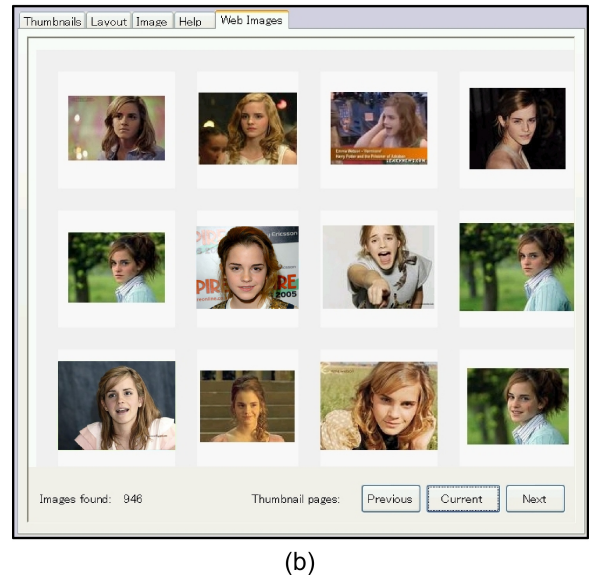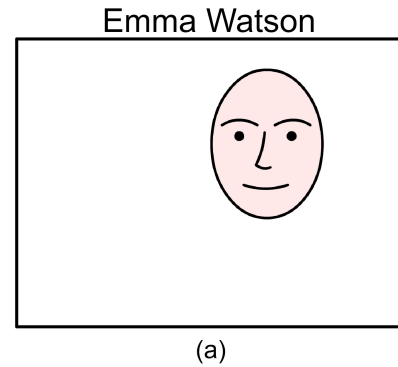
(a)

(b)

Figure 2: Using the interface for image retrieval.

5. CONCLUSION

We presented an interactive system for indexing editorial photographs using automatic face detection, and retrieving them using visual queries. Embedded comments remove the burden of re-indexing images upon transfer, and intensive image analysis during indexing. The iconic sketch allows a user to query for a wide range of compositions in a fast and intuitive manner. The iconic sketch can also be used to enhance personal photo organizers such as *Apple iPhoto* and *Picasa.* We are working on incorporating face recognition for faster indexing of images.

6. REFERENCES

[1] Haap media. stock.xchng - the leading free stock photography site. In *http://www.sxc.hu/,* 2001-2009.

[2] C. de Silva, T. Yamasaki, and K. Aizawa. Embedded tags and visual querying for interactive photo retrieval. In *Proceedings of PCM 2008*, December 2008.

[3] K. Rodden. *Evaluating similarity based visualizations as interfaces for image browsing.* Technical Report UCAM-CL-TR-543, University of Cambridge, 2002.

[4] S. Westman and P. Oittinen. Image retrieval by end-users and intermediaries in a journalistic work context. In *Proc. ICIIC 2006*, volume 176, pages 102–110, October 2006.