Organizing Photos of People

Andreas Girgensohn, John Adcock, and Lynn Wilcox
FX Palo Alto Laboratory
3400 Hillview Avenue
Palo Alto, CA 94304, USA
{andreasg, adcock, wilcox}@fxpal.com

ABSTRACT
As the size of the typical personal digital photo collection reaches well into the thousands or photos, advanced tools to manage these large collections are more and more necessary. In this demonstration, we present a semi-automatic approach that opportunistically takes advantage of the current state-of-the-art technology in face detection and recognition and combines it with user interface techniques to facilitate the task of labeling people in photos. We show how we use an accurate face detector to automatically extract faces from photos. Instead of having a less accurate face recognizer classify faces, we use it to sort faces by their similarity to a face model. We demonstrate our photo application that uses the extracted faces as UI proxies for actions on the underlying photos along with the sorting strategy to identify candidate faces for quick and easy face labeling.

Keywords: Digital photo collections, face detection and recognition, user interface design, home users.

INTRODUCTION
As digital cameras become increasingly ubiquitous, personal digital photo collections are growing both in number and size. A major challenge is to make photo organization simple and quick while enabling users to enjoy, present, and share their voluminous collections. Towards this end, we have opportunistically applied automation in our photo organizer [2]. We augment the application with user interfaces where the automation is not robust, thereby leveraging what automatic analysis can provide without adding unnecessary complexity or requiring excessive user intervention to compensate for imperfect automation technologies. One area where we have applied this approach is in the management of photos of people. Current state-of-the-art technologies can accurately detect faces but are less accurate at automatic face recognition. We demonstrate how the combination of user interface techniques and face detection and recognition technologies can identify photos with faces, extract the faces, and use the faces themselves as a user interface resource to label photos and the faces within them.

RELATED WORK
There are numerous applications for viewing and managing collections of digital photos. Some of those applications use the presence or absence of people in photos as a sorting criterion. The PhotoFinder system [5] provides stickies with people's names that can be attached to points in photos. PhotoFinder relies entirely on the user identifying and manually labeling each person by dragging a label onto the appropriate photos [5]. The FotoFile system [4] automatically assigns a name to an automatically detected face and then asks the user to confirm the assignment. This interaction style requires user confirmation of each automatically labeled face and is not well suited for labeling hundreds of faces. The MediaBrowser [1] uses automatically detected metadata such as the presence of faces to help users organize their photos. The authors describe the ability to recognize specific faces and to automatically attach names as desirable features in future versions. The system of Zhang et al. [6] automatically determines the most likely identities for each detected face and presents a list of candidate names for the user to choose. This approach still requires a user interaction for every face rather than our approach of simultaneously assigning groups of faces.

UIST '04, October 24--27, 2004, Santa Fe, New Mexico, USA
ACM 1-58113-962-4/04/0010

OUR PHOTO ORGANIZER
Our application [2] shows the entire collection of photos in a vertically scrollable light table (right pane in Figure 1a). A tree widget (left pane in Figure 1a) displays the photo categories with a top-level node for each category type. Tree nodes are drag-and-drop targets for selections of photos. Categories can be assigned by dragging a selection of photos onto a category node in the tree view, thereby applying that category to each photo in the selection.

The tree view can also be used to quickly navigate the light table to the photos tagged with a particular value. Left-clicking on a tree node scrolls the light table to the first photo with the selected value or tag and makes that photo the current selection. Additional operations on tree nodes are available by right-clicking on a tree node and selecting an operation from a pop-up menu. For example, by right-clicking on a tree node and choosing "Show only these photos" from the pop-up menu, only the photos below that node are shown in the light table.

USE OF FACES
Our demonstration shows that when new photos are imported into the application, the face detector, running automatically in the background, extracts the detected faces. Our photo application presents the extracted faces as thumbnails in one view of the light table. The face detector [3] that we use is highly accurate for photos with faces where both eyes and the nose are visible. It returns the actual eye positions such that the orientation of the face can be determined. We use the eye positions to produce the thumbnail by cropping the photo in an attractive fashion reminiscent of typical portrait photos. We align an ellipse to the connection between the eyes and center it on a point 20% of the eye dis-

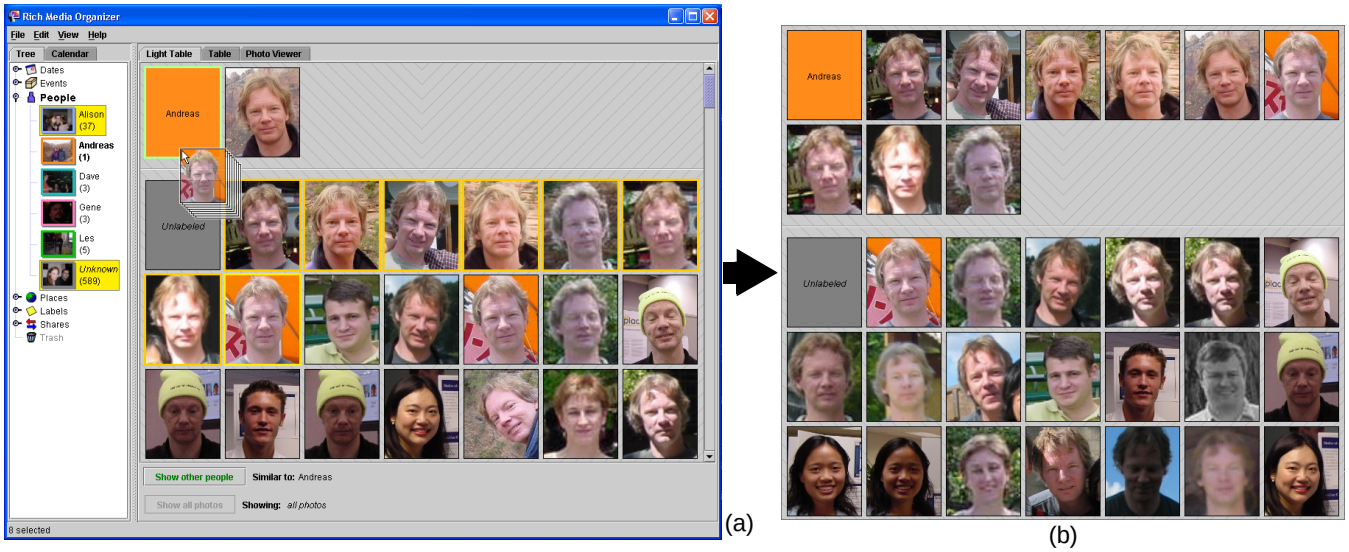

Figure 1: Finding pictures of Andreas. The faces are sorted by their similarity to the Andreas model. (a) The eight highest ranked selected photos (shown as a translucent stack) are being dropped on the Andreas icon to label them. (b) After dropping the faces. Note that the nine top-ranked faces can now be correctly assigned to the same person

tance below the eye connection. The rectangular bounding box of that ellipse determines the face crop.

The thumbnails can be manipulated just like photos (see Figure 1a). In particular, the user can use them as proxies for the photos when assigning the extracted faces to a person. A face can be assigned to a person simply by dragging it onto any of the iconic representations of that person in the tree or the light table (see Figure 1a). Once photos have been labeled with the people shown in them, they can be sorted or filtered by person so that photos depicting a single person or a combination of people can be found easily.

As the labeling of faces proceeds, the face recognizer forms a model for the person. This model is in turn used to sort the remaining unlabeled faces by similarity to the model. The sorted view allows the user to quickly decide which faces at the top of the similarity-sorted list depict the person of interest. Using a range selection, the user can select several faces and assign them to the appropriate person via drag-and-drop. Figure 1b shows the new view after the eight top-ranked faces were assigned to a person. Using the face recognizer to sort faces instead of having it automatically assign names to faces enables us to leverage on the strength of current face recognition technologies.

The model can be updated very quickly to incorporate the newly identified faces and the unlabeled faces are immediately resorted by their similarity to the new model. Also, models of hundreds of individual faces can be combined into a single model in a fraction of a second. Thus, updating models and similarity scores on the fly is practical.

CONCLUSIONS

We demonstrate how faces in photos can be a valuable resource to help with organizing large photo collections. While the face detection we employ has an extremely low false alarm rate, it is limited by only being able to find faces in frontal views. We expect as the technology matures, it will be able to detect faces in all aspects, sizes and angles. The techniques described here are incorporated in a commercial application that is currently in beta test. We will use the results of the beta test and other user feedback to further improve our photo organizer. Finally, we believe that our usage model can be adapted relatively easily for other face detection libraries that a) can return bounding boxes for the faces and optionally detect eye positions, b) can sort faces by similarity to a model, and c) can update the model quickly after more faces have been added. We welcome conference attendees to bring photos on removable media with which we can demonstrate our application.

REFERENCES

1. Drucker, S.M., Wong, C., Roseway, A., Glenner S., De Mar, S. MediaBrowser: Reclaiming the Shoebox. Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 433-436, 2004.
2. Girgensohn, A., Adcock, J., Cooper, M., Foote, J., and Wilcox, L. Simplifying the Management of Large Photo Collections. Human-Computer Interaction INTERACT '03, IOS Press, pp. 196-203, 2003.
3. Ioffe, S. Red Eye Detection with Machine Learning. IEEE International Conference on Image Processing, vol. II, pp. 871-874, 2003.
4. Kuchinsky, A., Pering, C., Creech, M.L., Freeze, D., Serra, B., and Gwizdka, J. FotoFile: A Consumer Multimedia Organization and Retrieval System. Proc. of the CHI 99 Conf. on Human Factors in Computing Systems, pp. 496-503, 1999.
5. Shneiderman, B. and Kang, H. Direct Annotation: A Drag and Drop Strategy for Labeling Photos. Proc. IEEE Intl. Conf. on Information Visualization, pp. 88-95. 2000.
6. Zhang, L., Chen, L., Li, M., and Zhang, H. Automated Annotation of Human Faces in Family Albums. Proc. of ACM Multimedia, pp. 355-358, 2003.