

# USER-ASSISTED PEOPLE SEARCH IN CONSUMER IMAGE COLLECTIONS

*Andrew C. Gallagher, Madirakshi Das, Alexander C. Loui*

Kodak Research Laboratories  
Eastman Kodak Company  
Rochester, New York 14650-1816  
alexander.loui@kodak.com

## ABSTRACT

In this paper, we investigate the process of searching for images of specified people in the consumer family photo domain. This domain is very different from the controlled environment of secure-access applications that have been extensively studied and face recognition packages are available in the market. Instead of the typical frontal mug shot, consumer photos are more likely to show people with unconstrained pose and illumination. This domain is also unique in that there are a large number of instances of a limited number of unique individuals. We develop and test facial recognition that is specifically targeted to this domain, using facial features that are derived from active shape modeling of faces, followed by a combination of the features using AdaBoost. We also provide a workflow that is suitable for lay users, and which rewards user inputs with improved performance. Test results show good performance on a challenging data set of consumer images.

## 1. INTRODUCTION

Studies on consumer picture-taking habits show that photographs containing people form the largest and most significant component of consumer photo collections. Because most people photographed by consumers are immediate family members, close friends, and relatives, a common set of people reoccur throughout their image collections. Associating images by the identifiable people (e.g. “these are pictures of my mother and sister”) is an intuitive way of searching for images in this domain. However, in the absence of manually assigned tags identifying the people present in images, retrieving photos containing particular persons continues to be a challenge. An automated indexing method that identifies the people in the images can provide this compelling search capability.

Current face recognition methods [1] are more suited to access control-type applications where a person looks directly at the camera, lighting is controlled, and there is not much variation in the appearance of the person. The U.S. Department of Defense conducted a Face Recognition Vendors Test (FRVT) in 2002 on the FERET data set [2]. While faces captured at different poses and with different illumination are included in FERET, the overall face

appearance variation is far smaller than that found in typical consumer images. Even then, the test showed that the best recognition performance degraded to 50% when outdoor images were used, and there were sharp declines in performance with younger subjects. In the family photo domain, a wide variation in lighting conditions (both indoor and outdoor) is expected, and babies and children are present in a significant proportion of images. Therefore, using a general-purpose face recognition solution on a standard adult face recognition database such as FERET does not provide meaningful performance measures in this domain. The workflow of providing registration mug shots of each unique individual before face recognition can be initiated is also not suitable for the consumer domain. Riya (riya.com) attempted an alternative workflow where the user could tag faces with names and confirm automatically identified faces. However, this workflow expects significant work from the user.

Our work investigates the characteristics of family photo collections, some of which can be exploited to improve recognition. Features and algorithms used in face recognition are also developed to be appropriate for this domain. As reported in a study on consumer attitudes towards automated image organization [6], users would like to have manual control over the algorithm-generated processes. This includes correcting wrong results, and expecting user input to have a positive effect on future system behavior. This paper describes a workflow supporting the ability to search for faces similar to a query face without the need for a registration phase or tagging, and allowing for user corrections and labels that provide performance improvements in the system.

## 2. CONSUMER FACE DATABASE

A consumer image database has been developed from personal image collections of 20 individuals. The owners labeled the people in their images to provide ground truth for testing. Labeled faces have both eyes clearly visible, which is a minor limitation because near-frontal faces are more relevant when searching for images of people.

Analysis of the collected labels provides a rich set of information useful for both understanding the behavior of

the photographer and for designing person recognition algorithms. Overall, the database contains 6475 images with 4711 labeled people; where 440 individuals are labeled more than once. There are over 30 examples for each of 34 individuals. A closer look at these 34 reveals that 25 are children. The conclusion that a large portion of the most popular individuals is babies or children aligns nicely with intuition. One may speculate that the frequently photographed individuals are important to the collection owners. However, it was found that some important people may appear rarely due to physical situations (e.g. a grandfather who visits only a few times each year.)

Statistics can be gathered relating the likelihood of an individual appearing in an image pair captured closely in time, using the database images that have valid date and time information in the file header. For example, if Person A is present in image  $i_1$ , it is probable that he is present in image  $i_2$  captured 1 s later. The capture time has great relevance toward determining who is in an image. Estimates of these likelihoods can be extracted from the person recognition database. For any two images  $i_1$  and  $i_2$ , with  $t_2$  being the difference in the image capture times of the two images, we are interested in computing the probability  $P(i_{2x}|(i_{1x}, t_2))$  of individual  $x$  being in image  $i_2$ , given that individual  $x$  is in image  $i_1$ . In Fig. 1, the solid line shows the value of  $P(i_{2x}|(i_{1x}, t_2))$  as a function of the image capture time difference  $t_2$ . As intuition would suggest, the probability decreases as the capture

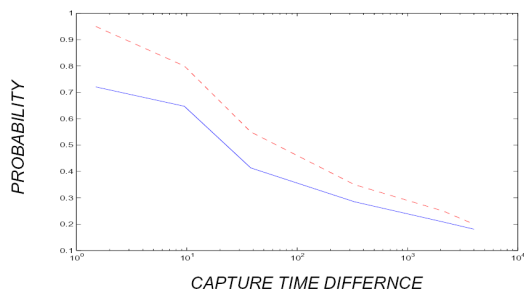


Fig. 1: Conditional probability distribution of an individual reappearing in another image.

time difference  $t_2$  increases. The dashed line in Fig. 1 is created by augmenting the actual data to account for the fact that sometimes individuals are in an image but not labeled because of pose, blur, resolution, or other problems. This information can be used to improve people detection results by deducing the presence of a person in an image based on his presence in another image taken very closely in time.

### 3. PEOPLE RECOGNITION

The first step in most recognition algorithms is obtaining a normalized face with eyes located at standard points. It is not possible to align faces to a standard template accurately

in our application because the eye locations are approximate, being automatically generated from the face detection output of a cascaded face detector described in [3]. Instead, we use an active shape model-based facial feature points finder [4] to localize the face more accurately. The starting condition of the points finder is initialized by using the approximate eye locations provided by the face detector.

#### 3.1 Facial features

Facial features used for this application need to be robust to some pose variations around the frontal pose. They also need to handle illumination variation that appearance-based methods are sensitive to.

A compact representation of facial geometry can be obtained by determining the principal components of the facial feature points by first gathering the feature point locations from a large number of images of people. Then each principal component describes a variation of a particular set of facial feature points from the average set of facial feature points. A good set of features, called Eigen-appearance features, is obtained by projecting a set of feature points onto the principal components that relate to the differences in appearance between unique individuals, and ignoring the principal components that relate to changes in expression or pose. For example, it was observed (see Fig. 2) that the 0<sup>th</sup> and 2<sup>nd</sup> principal components computed relate to pose (side-to-side and up-down motion of the head, respectively), and are therefore ignored as irrelevant to face recognition, while the 1<sup>st</sup> and 3<sup>rd</sup> principal components relate to variations in appearance and are useful for recognizing faces. The selected features display some insensitivity to pose and expression that is crucial for recognizing faces in this domain.

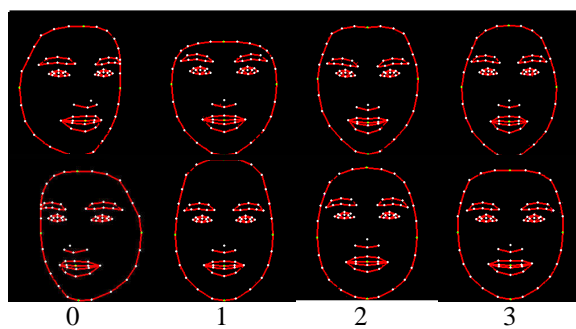


Fig. 2: The variation of facial geometry along the first four principal components is shown.

#### 3.2 Computing facial similarity score

For each feature, the distribution of difference in value is computed for two classes—each person to same person comparisons and each person to different person comparison—using the face recognition database ground

truth. Simple Bayesian threshold-based binary classifiers are constructed for each feature based on these distributions. The AdaBoost algorithm (as in [3]) is used to determine the weights to be assigned to the weak single-feature classifier outputs so that features are weighted based on their discrimination power. A distance score between any two faces is computed by using a linear combination of the raw feature differences, weighted by the AdaBoost weights.

### 3.3 Classification network

For classification, a probability network (see Fig. 3) is formed by establishing links between each point's  $k$  nearest neighbors. Each link has an associated probability weight  $w_{ij}$  (where  $i$  and  $j$  represent the  $i^{\text{th}}$  and  $j^{\text{th}}$  points joined by the link). The weight describes the probability that points  $i$  and  $j$  have the same identity. The weights are based on the distance between points  $i$  and  $j$ , computed as described above.

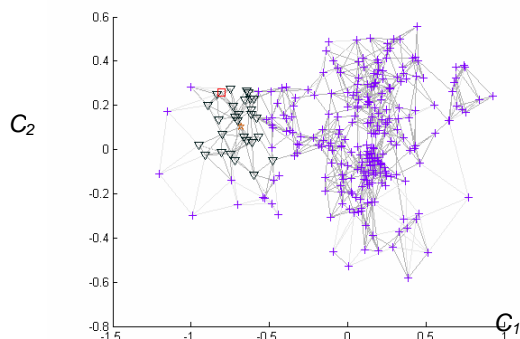


Fig. 3: An example of the propagation of beliefs across the network. The star shows a point with an assigned label (i.e. 100% likelihood that the identity of that person is q1.) The network then shows all the points (29 triangles (true matches) and 1 square (false match)) that, after propagation, have a greater than 50% probability of being individual q1.

Some of the points in the network may have associated labels (gathered through a user interface) that can be propagated through the network as follows. If the identity of the  $i^{\text{th}}$  person is known to be  $q$ , then that belief is propagated iteratively through the network according to the weights  $w_{ij}$ . The  $j^{\text{th}}$  feature point then has belief of  $w_{ij}$  that its identity is  $q$ . In a subsequent iteration, this belief can be propagated to  $j$ 's nearest neighbors. When multiple beliefs associated with the same individual arrive at a single point, only the maximum value belief is preserved.

### 3.4 People recognition interface

Figure 4 shows an interface for enabling search by people present in consumer image collections. The user can bring any image into the active image area and add face(s) in the image to the search box. Multiple faces from different

images can also be added to the search box. In this case, the search retrieves images containing matches to all the faces in the search box. The user is also presented with a set of icon faces that represent the clusters of similar people [5] detected in the collection. Attaching labels to the icon faces can label the whole cluster with a name. Search labels and/or icon face(s) can also be added to the search box, eliminating the need for first finding images that contain a particular individual.

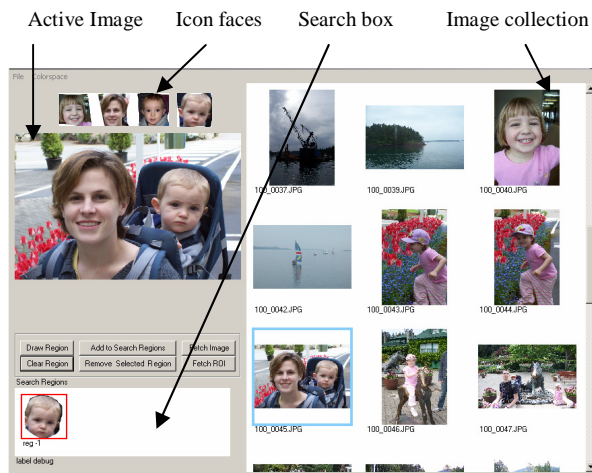


Fig. 4: Interface for querying a collection for finding images of specified people.

## 4. RESULTS

The person recognition database was used to evaluate the performance of face recognition in the following scenario. The user identifies a person of interest by selecting a single face from an image in the image collection (see Fig. 7) and instructs the system to “find images containing this person.” The classifier then determines, for each image in the collection, the likelihood that the image contains the person of interest. The accuracy of the classifier is quantified by classifying images with a computed likelihood greater than a threshold  $T_0$  as containing the person of interest. Figure 5 shows an ROC curve generated by computing the precision and the recall while varying  $T_0$ . The thick black trace summarizes the performance of the system, averaged over all four popular individuals. Each ROC curve has one point that represents the performance of the system while displaying the first 25 image results from the search. On average (i.e. randomly selecting a face and searching for that person), the system produces about 21.3 images (85% precision) containing the person of interest and 3.7 images that do not contain the person of interest. As a benchmark, random selection would result in only 40% precision.

When searching for a particular person, the classification results vary based on the specific instance of the person selected. To take this into account, classification

is initiated multiple times (once for each appearance of the person). The performances from each of these searches are combined to produce the overall performance. Over all the popular people in the face recognition database, the precision is 53% (compared to 29% as the random baseline) for the first 25 image search results. This precision score uses only a single labeled example for recognition, and includes the effects resulting from face detector misses.

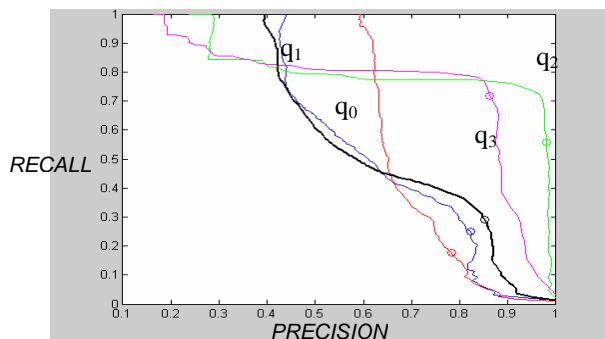


Fig. 5: Classification performance for the four most popular individuals from one image collection of the person recognition database.

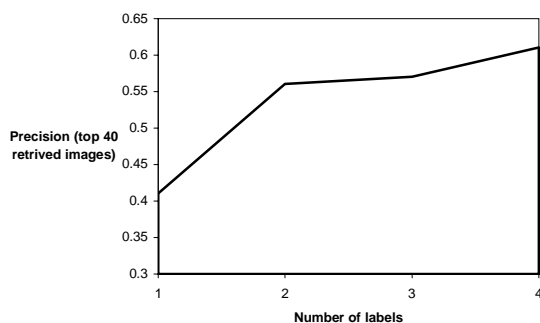


Fig. 6: Improvement in precision as the user labels examples of an individual.

In addition to simple search by example, our system allows the user to correct the labels of wrongly identified individuals, as well as label example faces for each frequently occurring individual in the collection. Since these labels are propagated in the probability network, correcting a few wrongly labeled faces or providing correspondence between faces that are not closely linked in the original network, results in improved results. Figure 6 shows an example of improvements in precision as a result of the user labeling a small number of example faces of an individual. This is particularly valuable when the search performance is unsatisfactory with no labeled faces. For example, there may be strong facial resemblance between people in the family data set that the feature set cannot disambiguate.

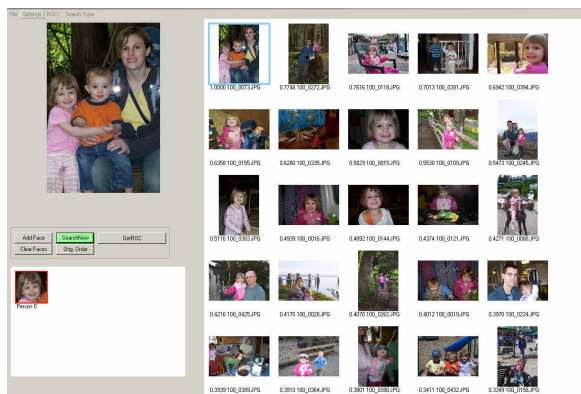


Fig. 7: Search results after querying a collection for images of a specified person.

## 5. CONCLUSIONS AND FUTURE WORK

A face recognition database has been compiled using images from 20 personal image collections. The database is used to develop effective features for person recognition in consumer image collections. Features related to facial geometry with tolerance for expression and pose are used to describe facial appearance. In addition, the time between image captures is used. A classifier is developed by forming a probability network between faces having similar features. A suitable workflow for the consumer domain is proposed. Performance of the system for finding images containing a query face is quite good, with further improvements as a user labels faces. This technique appears promising for managing consumer image collections without forcing the user to undergo an extensive manual training stage.

Future work includes expanding the interface to incorporate other types of semantic features (e.g. events and locations) to create a comprehensive search framework for the consumer image domain.

## REFERENCES

- [1] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, Vol. 35, No. 4, pp. 399-458, December 2003.
- [2] P. Phillips et al., "Face recognition vendor test 2002: evaluation report," NISTIR 6965, 2003. Available online at <http://www.frvt.org>.
- [3] P. Viola and M. Jones, "Robust real-time face detection," *Int. Journal of Computer Vision*, Vol. 57, No. 2, pp. 137-154, 2004.
- [4] M. Bolin and S. Chen, "An automatic facial feature finding system for portrait images," *Proc. of IS&T PICS*, pp. 226-231, 2002.
- [5] M. Das and A. Loui, "Automatic face-based image grouping for albuming", *IEEE Int. Conf. on Systems, Man and Cybernetics*, Oct 2003.
- [6] C. Cerosaletti et al, "Approaches to Consumer Image Organization based on Semantic Categories", *Proc. of SPIE Optics East conference*, October 2006.