USING A MARKOV NETWORK TO RECOGNIZE PEOPLE IN CONSUMER IMAGES

Andrew C. Gallagher[1,2] *and Tsuhan Chen*[1]

[1]Carnegie Mellon University, [2]Eastman Kodak Company

ABSTRACT

Markov networks are an effective tool for the difficult but important problem of recognizing people in consumer image collections. Given a small set of labeled faces, we seek to recognize the other faces in an image collection. The constraints of the problem are exploited when forming the Markov network edge potentials. Inference is also used to suggest faces for the user to label, minimizing the work on the part of the user. In one test set containing 4 individuals, an 86% recognition rate is achieved with only 3 labeled examples.

Index Terms— face recognition, Markov network

1. INTRODUCTION

Studies on consumer image collections show that images containing people form a very significant component, and most images contain one or more people. Since most people photographed by consumers are immediate family members, close friends or relatives, a common set of people re-occur throughout their image collections. Fig. 1 shows a few image examples from an image collection. Labeling images by the identifiable people (e.g., these are pictures of my mother and sister) allows the collection to be searched. However, labeling is a very labor-intensive process. In the absence of manually assigned labels, retrieving photos of particular persons is a challenge. The goal of this work is to identify people in consumer images, thus enabling simple retrieval. At first, none of the faces in the images are labeled, though we do make the simplifying closed world assumption [1] that we know the number of people present in the image collection. We expect that with very few labeled face examples (e.g. 1 example) that the system can begin to properly identify other faces in the image collection. As more faces are labeled by the user the performance will improve.

Certainly, there are many techniques for recognizing faces, or for comparing the similarity of two faces [2]. However, there are significant differences between face recognition in general and the problem of recognizing people in consumer images. The field of face recognition emphasizes the discovery of features that are useful for recognition, and generally ignores issues related to multiple people in a single image. Researchers are beginning to focus on the problem of recognizing faces in consumer image collections. For example, a nearest neighbor classifier is used for face annotation [3].

Fig. 1. Example of a few images from an image collection.

We propose using a Markov network based on the natural constraints of the problem for recognizing the faces in the image collection. The *unique object constraint* states that since an individual can only appear once in an image (barring mirrors or images of images), any faces from a single image must be different individuals. This Markov network provides superior performance over a nearest neighbor classifier. Further, the network actively determines which face the user should label next to provide the most information for labeling unreconized faces.

1.1. The Consumer Face Database

A database of consumer image collections was developed to explore this problem. Eight image collections were acquired, containing a total of 1084 images of people (an average of 136 images per collection.) The database includes a total of 1952 labeled instances of 165 unique people. Analysis of the collected face identities provides a rich set of information for developing recognition algorithms. Overall about 75% of all images contain one or more people, and of these, nearly half contain more than one person. About 14% of the individuals appear in greater than 15% of their collection images. These popular people are the ones we would like to be able to automatically identify, as they are obviously important to the photographer. In our eight image collections, the number of popular people ranges from one to five.

2. IMAGES AND FEATURES

A face detection algorithm [4] is used to detect faces in each image. Next, an active shape model [5] is used to locate the

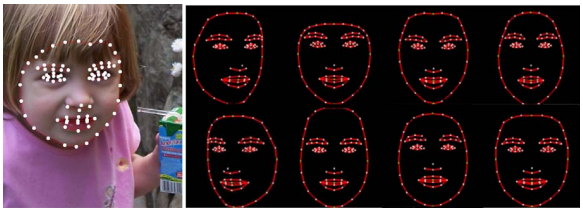

Fig. 2. **Left:** An image with 82 key points automatically extracted. **Right:** PCA is used to represent each face with a 5-dimensional feature vector, corresponding to eigenvectors that relate to differences in individual appearance. A visualization of the first four eigenvectors of the key points is shown. The top row corresponds to the average face plus the eigenvector, and in the bottom row the eigenvector is subtracted from the average face. The first and third eigenvectors relate to facial pose and are ignored. The second and fourth (and three other) eigenvectors relate to differences in individual appearance and are used in our experiments.

positions of 82 key points for each face. Facial features based on facial geometry are robust to some pose variations and illumination variation that is typically encountered in consumer photography. An example face having automatically located key points is shown in Fig. 2. The feature vector displays some insensitivity to pose, illumination, and expression that is crucial for recognizing faces in this domain. The feature vectors associated with faces from an image collection can be visualized by plotting each face according to the first two dimensions of the feature space, as shown in Fig. 3.

3. BUILDING A PAIRWISE MARKOV NETWORK FOR INFERENCE

The identity of each face is considered to be a random variable X_n that can take on values corresponding to each individual. For the first image collection, the joint probability distribution is $P(X_1, \ldots, X_n, \ldots, X_N)$ where each X_n can take on values in the set Vals$(X_n) = \mathbf{p} =$\{Hannah, Jonah, Andy, Holly\}. Given that some of the X_n's are observed, the goal of classification is to determine the most likely assignment (MAP) of the unobserved variables. A pairwise Markov network is formed over the faces. The formation of the network is based on the observation that faces close in feature space are likely to be the same individual, and by a constraint we call the *unique object constraint*. The unique object constraint states that since an individual can only appear once in an image (barring mirrors or images of images), any faces from a single image must be different individuals. The edge potentials of the Markov network are created by the following rules:

1. A *dissimilarity edge* is formed between X_i and X_j if X_i and X_j are faces from the same image.

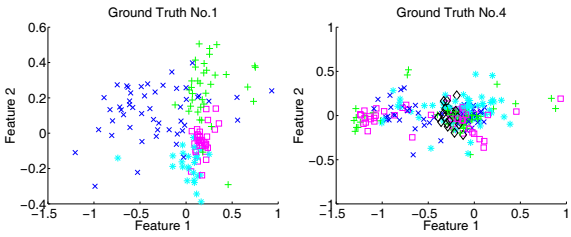

Fig. 3. A visualization in feature space of the faces from two image collections. Each individual's feature vectors are plotted with a different symbol. The two image collections contain 146 and 261 faces, with 4 and 5 unique individuals respectively.

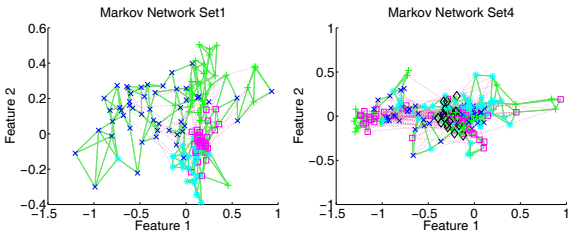

Fig. 4. The Markov networks for two image collections. Similarity edges are green solid lines, and dissimilarity edges are dashed magenta.

2. A *similarity edge* is formed between X_i and X_j if X_i is one of the M closest faces to X_j in feature space (measured by Euclidean distance.)

The Markov networks (with M = 3) for the image collections from Fig. 3 are shown in Fig. 4.

The potentials for the two types of edges must be defined. Each potential factor is a matrix of size K x K, where K is the number of unique individuals in the image collection. A dissimilarity edge has a potential function $\Psi_D(x_i, x_j)$ with small values on the diagonal:

$$\Psi_D(x_i, x_j) = \exp(-\beta\delta(x_i, x_j)) \quad (1)$$

where $\delta(x_i, x_j)$ is an indicator function that is zero except when x_i equals x_j.

The similarity potential functions are related to the Euclidean distance $D(X_i, X_j)$ between X_i and X_j in feature space. This potential function can be learned from data by selecting many pairs of faces from several training image collections. Fig. 5 shows the learned probability that $x_i = x_j$ given the distance $D(X_i, X_j)$, and an exponential approximation to this probability. Notice that for distances $D(X_i, X_j)$ greater than a certain value (≈0.4), the probability essentially becomes the prior probability $\frac{1}{K}$. The similarity potential

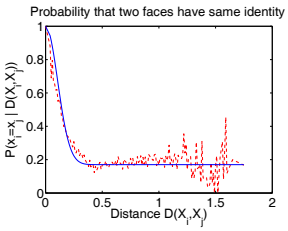

Fig. 5. The probability that two faces are of the same individual, given the distance in feature space. This relationship is used to establish the similarity potentials. The solid exponential curve is fit to the dashed data curve.

$\Psi_S(x_i, x_j)$ is defined as:

$$\Psi_S(x_i, x_j) = \begin{cases} \frac{1}{K} + \frac{K-1}{K} \exp\left(-\gamma D(X_i, X_j)^2\right), & x_i = x_j \\ \frac{1}{K} - \frac{1}{K} \exp\left(-\gamma D(X_i, X_j)^2\right), & x_i \neq x_j \end{cases} \quad (2)$$

4. EXPERIMENTAL RESULTS

The Markov network defines a joint probability distribution over the nodes (the identities of faces). Given this joint probability distribution, marginalization can be used to answer queries. However, computing this joint probability distribution is generally intractable and approximate inference techniques must be used. The evidence (faces with known identity) is considered and inference is performed with loopy belief propagation (LBP) [6]. In our work, $\gamma = 44$ and $\beta = 3.9$.

The following experiment is performed to simulate the effect of labeling faces in the image collection. A random ordering of the faces is established. Faces are labeled according to the order and inference is used to classify the identity of all remaining unlabeled faces. Each classification is compared against the true label to find the classification rate. This experiment is repeated for 10 random orderings. Results for networks created with $M = 3$ are shown in Fig. 6. For comparison, the Markov Network performance is compared with using a nearest neighbor classifier for classifying the identities of the unknown faces. LBP achieves an outstanding 83% correct classification rate on collection 1 after only 9 faces are labeled and shows steady improvement as the number of labeled faces increases.

4.1. Actively Selecting which Face to Label

In addition to wanting to search their image collections, users also desire to label as few faces as possible. Thus, we seek to identify the face, that when observed, will be the most helpful in solidifying the marginal beliefs for the remaining unlabeled faces. In other words, we want to identify the face, that when observed, provides the greatest reduction in the entropy of

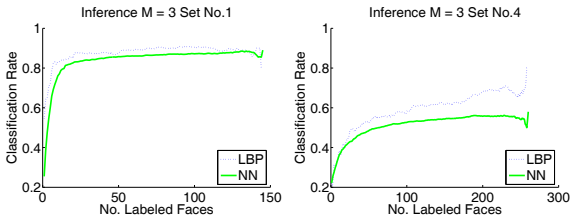

Fig. 6. The inference results for classifying face identity on two image collections as a function of the number of randomly selected labeled faces. The Markov network outperforms the nearest neighbor classifier.

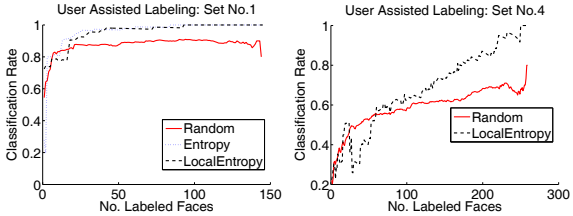

Fig. 7. The performance of inference with different orderings for labeling the faces. The selection order is determined either randomly (Random), by considering the mean field entropy (Entropy), or by using a heuristic that finds local regions of the network with high entropy (LocalEntropy).

the joint distribution of the remaining variables. This is an example of active learning [7, 8].

More formally, if the set of observed (labeled) faces is $\mathbf{X}_o$ and the set of unlabeled faces is $\mathbf{X}_u$, then the joint distribution described by the Markov network is $P(\mathbf{X}_u|\mathbf{X}_o)$. We seek the identity of the face X_i, that when observed produces the distribution with the minimum expected entropy $H(\mathbf{X}_u - X_i|\mathbf{X}_o \cup X_i)$. The choice for the next face for the user to label X_H is:

$$X_H = \arg\min_{X_i} E_{X_i}\left[\hat{H}(\mathbf{X}_u - X_i|\mathbf{X}_o \cup X_i)\right] \quad (3)$$

$$= \arg\min_{X_i} \sum_p \hat{P}(X_i = p)\hat{H}(\mathbf{X}_u - X_i|\mathbf{X}_o \cup X_i) \quad (4)$$

$$\approx \arg\min_{X_i} \sum_p \hat{P}(X_i = p) \sum_{X_j \in \mathbf{X}_u - X_i} \hat{H}(X_j|\mathbf{X}_o \cup X_i) \quad (5)$$

where $\hat{P}(X_i = p)$ is the current belief that face X_i is a particular individual p from the set $\mathbf{p}$. Calculating the entropy $\hat{H}(\mathbf{X}_u - X_i|\mathbf{X}_o \cup X_i)$, the entropy of the joint distribution from approximate inference, is computationally intractable, so we proceed from (4) to (5) using the mean field approximation that each variable is independent. This is estimated by performing approximate inference, then computing the en-

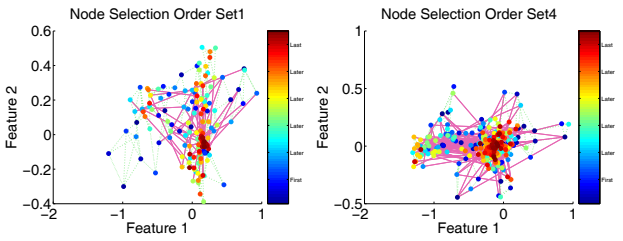

Fig. 8. The order of of labeling faces in the Markov network, based on the LocalEntropy heuristic.

	Hannah	Jonah	Holly	Andy
Hannah	33	15	0	1
Jonah	1	32	1	1
Holly	1	0	24	0
Andy	0	0	0	34

Table 1. Confusion matrix for inference results on Set 1 with only three labeled faces.

tropy of the resulting node beliefs. The minimization in (5) requires that each variable be considered at each value that it can take on, so inference must be performed a total of $K * N$ times for selecting the next face for the user to label. Obviously, time can be a problem. For example, in Fig. 8 set 4 does not have a performance for the (Entropy) ordering as it would cost nearly one week of computing time!

A heuristic (LocalEntropy) is used to greatly reduce the computational burden. Rather than considering the effect of observing a given X_i over all the unlabeled faces, we look for local regions of the Markov network with high entropy. Observing a variable in a given local region is sure to decrease the local entropy, so the face associated with the highest local entropy is posed to the user to be labeled. With this heuristic, the choice for the next face for the user to label X_L is:

$$X_L = \arg\max_{X_i} \left[2\hat{H}(X_i) + \sum_{X_j \in \text{Nei}(X_i)} \hat{H}(X_j) \right] \quad (6)$$

The local entropy heuristic produces good results and is easy to compute, as no inference is required in its computation.

Fig. 7 compares the performances of the difference methods for selecting the next face to label. Fig. 8 shows the recommended order to provide labels for the faces using the heuristic from the first faces to label (dark blue) to the last (red). As expected, the first few faces to observe are widely scattered throughout the network, intuitively providing a distribution of information throughout the graphical network.

Set 1 provides an interesting case study of using the expected entropy to select which faces to label. Using that method, after labeling the first three faces, LBP inference on the network achieves a remarkable 123 of 143 correct identifications, a rate of 86%. This is all the more impressive considering that

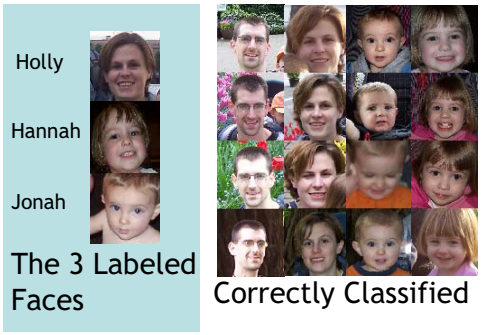

Fig. 9. **Left**: The first three faces that are labeled according to (Entropy). **Right**: Correctly recognized faces even though there were no labeled examples of Andy.

this collection contains four individuals, so one individual is recognized well even without a single training sample! Fig. 9 shows the three labeled faces and some faces that were correctly recognized, and Table 1 contains the confusion matrix.

5. CONCLUSIONS

A Markov network provides effective face recognition in consumer image collections. The Markov network outperforms nearest neighbor classification. In addition, the Markov network actively recommends faces to be labeled by the user. In the future, we plan to extend the work to include a node potential for each face based on the facial features, where the parameters are learned from the labeled faces.

Acknowledgement: The authors wish to thank Carlos Guestrin for his enlightening advice on this work.

6. REFERENCES

[1] S. Russell, "Identity uncertainty," in *Proc. IFSA-01*, 2001.

[2] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, 2003.

[3] L. Chen, B. Hu, L. Zhang, M. Li, and H. Zhang, "Face annotation for family photo album management," *Int. Jour. of Image and Graphics*, 2003.

[4] M. Jones and P. Viola, "Fast multiview face detector," in *Proc. CVPR*, 2003.

[5] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models-their training and application," *CVIU*, 1995.

[6] T. Meltzer and Y. Weiss, "c_inference," Downloaded Nov. 2006 from Hebrew University. http://www.cs.huji.ac.il/~talyam.

[7] L. van der Gaag and M. Wessels, "Selective evidence gathering for diagnostic belief networks," *AISB Quarterly*, 1993.

[8] C. Zhang and T. Chen, "Annotating retrieval database with active learning," in *Proc. ICIP*, 2003.