Using Group Prior to Identify People in Consumer Images

Andrew C. Gallagher
Carnegie Mellon University
Pittsburgh, Pennsylvania
agallagh@cmu.edu

Tsuhan Chen
Carnegie Mellon University
Pittsburgh, Pennsylvania
tsuhan@cmu.edu

Abstract

While face recognition techniques have rapidly advanced in the last few years, most of the work is in the domain of security applications. For consumer imaging applications, person recognition is an important tool that is useful for searching and retrieving images from a personal image collection. It has been shown that when recognizing a single person in an image, a maximum likelihood classifier requires the prior probability for each candidate individual. In this paper, we extend this idea and describe the benefits of using a group prior for identifying people in consumer images with multiple people. The group prior describes the probability of a group of individuals appearing together in an image.

In our application, we have a subset of ambiguously labeled images for a consumer image collection, where we seek to identify all of the people in the collection. We describe a simple algorithm for resolving the ambiguous labels. We show that despite errors in resolving ambiguous labels, useful classifiers can be trained with the resolved labels. Recognition performance is further improved with a group prior learned from the ambiguous labels. In summary, by modeling the relationships between the people with the group prior, we improve classification performance.

1. Introduction

Figure 1 shows a few example images containing people from a single image collection. Because each person is a unique individual, we immediately have a powerful constraint that affects the design and selection of a classifier. Within an image, an individual can appear at most one time, and each person in an image can be only one individual [1]. (We ignore the rare images containing a face and its mirror reflection, or images containing other images, etc.) This intuitive constraint provides a foundation for determining the identities of people from consumer images. We call this constraint the *unique object constraint*.

Figure 1. Example of a few images from an image collection. Ambiguous labels provide the information about who is in each image and are used to estimate the group prior.

When multiple people are in an image, there is usually a relationship between the people in the image. For example, the people could be friends, co-workers, siblings, or relatives. By learning the prior probability of different individuals appearing together in an image, classification can be improved. This prior probability of certain groups of people appearing in an image is called the *group prior*. The group prior implicitly incorporates the unique object constraint, because the probability of any person appearing more than once in an image is zero.

Ambiguous labels are sometimes supplied with a set of images containing people. An ambiguous label provides a label for a unique object that appears in an image, without indicating which object is associated with which label. Figure 1 shows the ambiguous labels associated with several images. Ambiguous labels for individuals' names in images occur naturally in several situations. First, many software packages (e.g. www.flickr.com) allow the user to tag images with any keyword related to the image. Second, many people annotate their images with captions such as "George and Martha in their canoe" which conveys that Martha and George are in the image but does not indicate which is George and which is Martha. We seek to resolve the ambiguous labels by assigning each label to a specific face in the image. In addition, ambiguous labels provide exactly the information we need to estimate the group prior, which can be used to improve classification performance.

In this paper, we present algorithms that incorporate the group prior to model the relationships between people in the images. In Section 2, we review the related work. We describe a database for recognizing people in consumer images in Section 3. We then describe an algorithm to resolve ambiguous labels (Section 4). Finally, in Section 5, we show how labeling a small image set with ambiguous labels can be used to learn group prior information and train classifiers that recognize faces in previously unseen and unlabeled images for the purpose of automatic annotation or retrieval.

2. Related Work

Certainly, there are many techniques for recognizing faces, or for comparing the similarity of two faces [14]. However, there are many significant differences between the problem of face recognition in general and the problem of recognizing people in consumer images. The field of face recognition emphasizes the development of features that are useful for recognition, and generally ignores issues related to prior probabilities (of an individual or specific group of individuals appearing in an image.)

With regard to capitalizing on problem-specific constraints, several classification and clustering algorithms have been developed that either implicitly or explicitly examine constraints to improve the performance of the classifier. In unsupervised clustering, Wagstaff *et al.* [11], describe an algorithm that uses known constraints between example points. The "must-link" constraint requires that two examples be in the same cluster while the "cannot-link" constraint requires that the two points cannot be in the same cluster. Constraints have also been added to clustering algorithms such as normalized cut [9, 7]. The constraints can relate to lane segmentation [11], image segmentation [12], or inferring web page relationships [9]. When considering faces from many images, all faces from a single image are all mutually "cannot-link" due to the unique object constraint and there are no "must-link" constraints. These approaches do consider the problem constraints, but they do not incorporate labeled data and are not suitable for our application.

Computer vision researchers have worked with ambiguously labeled data. Satoh and Kanade [8] developed the "Name-It" system to identify faces in news video from the transcripts and video captions. Berg *et al.* [1] extract names from captions of news photos and associate the names with faces in the images. Both these applications involve noisy labels (i.e. a detected name may not be someone who appears in the image) and are difficult problems. Berg handles this noise by initializing the name-face assigment algorithm using those images containing only a single face with only one name in the caption, then uses expectation maximization to assign names to faces. Our ambiguously labeled images are related to this work, but we assume that a human is actively providing the ambiguous labels for each image's detected faces. Thus, we expect that a name provided by the human will appear in the image, and therefore avoid the noisy label problem.

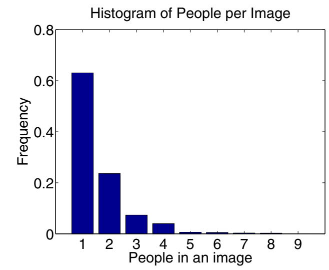

Figure 2. **Left:** A histogram of the number of people per image from a set of four image collections of over 3500 consumer images.

In an example of using weakly labeled data, Zhang *et al.* [13] describe a photo organizing system where a user indicates a set of images that contain a certain person, and the system selects one face from each of the images that maximize the overall similarity between the selected faces.

Our work builds on these techniques by improving the recognition performance using a group prior. The group prior serves as the context for the classification problem, akin to performing object detection by setting the context of the scene [10]. The cooccurance of individuals in images has been considered by Naaman *et al.* [6] for an interactive image labeling application that uses only image context (like the image capture time and place, and other people in the image) to suggest the next most likely label name for an image. We build on the work of Naaman *et al.* by finding the prior for any group or people (rather than single person) in the image, and combining that prior with facial features. Our work extends that of Zhang *et al.* by simultaneously handling multiple person names to disambiguate the ambiguous labels. Our ambiguous label resolution algorithm handles a simpler problem than either [1, 8] yet it does not need to be initialized with faces having known labels. In summary, classification is improved by considering the features of all people in the image along with the group prior.

3. Images and Features

Much of the work described in this paper takes advantage of constraints that naturally occur when multiple persons appear in a single image. Therefore, it is important to understand the distribution of people in images.

Four image collections were acquired, containing a total of 1084 images with people. Each collection owner labeled the people in each image. The database includes 1924 labeled instances of 114 unique people. Analysis of the col-

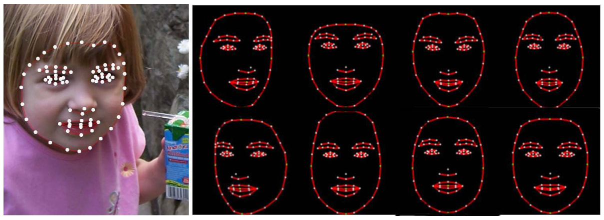

Figure 3. **Left:** An image with 82 key points automatically identified. **Right:** PCA is used to represent each face with a 5-dimensional feature vector, corresponding to eigenvectors that relate to differences in individual appearance. The visualization of the first four eigenvectors of the key points 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 eigenvectors relate to differences in individual appearance and are preserved.

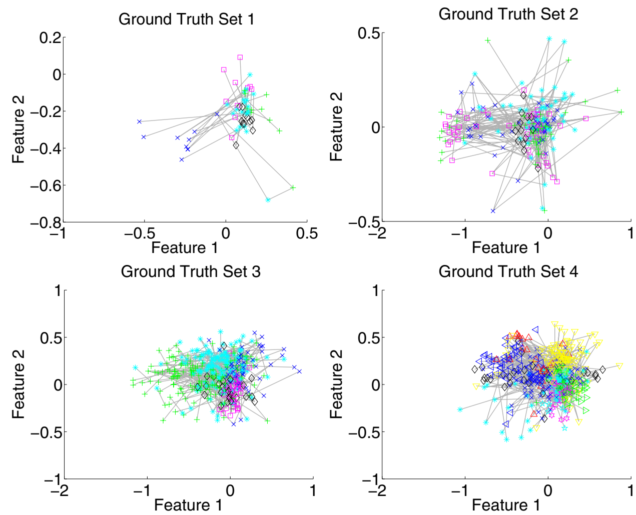

Figure 4. The four test image collections. Each data point represents a face (projected to the first two feature dimensions). Each unique symbol represents a different individual in that image collections. Lines connect faces that appear in the same image.

	Set 1	Set 2	Set 3	Set 4
Total images	300	300	1197	2099
Images with multiple people	26	67	188	191
No. faces from these images	61	204	420	455
Constraints	44	237	288	360
Popular unique individuals	5	5	5	11

Table 1. Information about the four datasets.

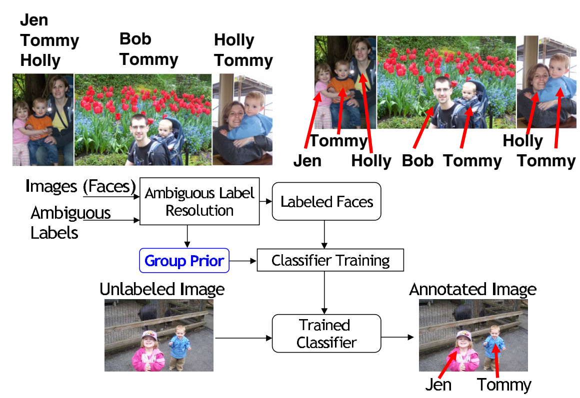

Figure 5. A system diagram. An image collection with ambiguous labels first has the ambiguous labels resolved. Then, a classifier is trained for each individual and the group prior is learned. Finally, faces in unlabeled images are classified.

lected face identities provides a rich set of information for recognition algorithm development. Figure 2 shows a histogram of the number of people in an image in images with people. About 50% of the images containes one or more people, and of these many contain more than one person. Each image collection has a small number of people that appear very often. These popular people are the ones we would like to be able to recognize, as they are obviously important to the photographer. In our image collections, the number of popular people ranges from five to eleven.

A face detection algorithm [3] is used to detect faces in each image. Facial features based on facial geometry are robust to some variation in pose and illumination that is typically encountered in consumer photography [14]. An active shape model [2] is used to locate the positions of 82 key points for each face, and each face is represented as a 5-dimensional feature vector. An example face having the automatically determined key points is shown in Figure 3. These features are not the state-of-the-art features for recognizing faces, but are sufficient to demonstrate our approach.

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 Figure 4. Each individual's feature vectors are plotted with a different symbol. We are interested in studying the group prior with images containing more than one of the popular unique individuals. The four image collections contain 61, 204, 420, and 455 faces with at least two faces per image, and 5, 5, 5 and 11 popular unique individuals respectively. In Figure 4, a line is drawn between faces that appear in the same image. This corresponds with the unique object constraint that since an individual can only appear once in an image, any two faces joined by a line must be different individuals. The image collections have 44, 237, 288, and 360 total constraints, respectively. Each constraint is related to a unique pair of faces in an image. Table 1 summarizes these datasets which are used throughout this paper.

4. Resolving Ambiguous Labels

An ambiguously labeled image has associated individual names but the labels do not indicate which person is which individual. The caption of Figure 1 gives an example of ambiguous labels for three images. Once the ambiguous labels have been resolved, we have a collection of labeled faces. A classifier can be trained with these labels so that faces from completely unlabeled images from the same collection can be recognized. Figure 5 illustrates the proposed system.

We resolve the ambiguous labels by assigning each label to a person in an image. The objective function is the sum of squared distances between each face and the associated cluster center for its label. Certainly, minimizing this objective function by computing it for every possible assignment of labels is out of the question for all but the smallest number of faces and images.

Given a set of J ambiguously labeled images, the goal is to assign each face to a cluster C_k corresponding to one of K label names in the name set $\mathbf{N}$ (where K is the number of unique names among the ambiguous labels.) Let f_{mj} represent the features for the m^{th} face from the j^{th} image. M_j is the number of faces in the j^{th} image. Every image with more than one face has a unique object constraint that f_{mj} and f_{nj} cannot belong to the same cluster C_k, $\forall m \neq n$. An element $n^k \in \mathbf{N}$ is a particular name in the set. The notation n_m^k indicates that the name n^k is associated with person m from an image. In addition to the unique object constraint, we have an additional constraint that each image's faces can only be assigned to a subset of the possible labels $\mathbf{N}$ (the ambiguous labels for that image). For image j, the ambiguous labels are $\Psi_j \subseteq \mathbf{N}$.

An algorithm for resolving ambiguous labels is ALR:
ALR: Ambiguous Label Resolution Algorithm

1. For each image j, randomly assign faces f_{mj} to ambiguous labels Ψ_j.

2. Compute the parameters of each label's cluster from the faces assigned to that label.

3. For each image j, assign faces f_{mj} to labels Ψ_j in a manner that respects the unique object constraints and minimizes the overall squared distance E_j for the image, using the Hungarian algorithm [4].

4. Iterate between 2 and 3 until convergence.

5. Return the final assignments of faces to clusters.

Step 3 requires further explanation. For each image j, we assign all faces from that image to ambiguous labels Ψ_j such that the sum of squared distances between each face and the corresponding cluster center C_k is minimized. We construct the matrix D, having elements d_{mk} where d_{mk} is the squared distance from the m^{th} face to the k^{th} cluster center, and $k \in \Psi_j$. Then, the Hungarian algorithm is used to find the optimal assignment of faces to clusters (in polynomial time) that minimizes the overall squared distance E_j for the image. The residual error for the j^{th} image is $E_j = \sum_{m,k}^{M_j} z_{mk} d_{mk}$, where z_{mk} is an indicator variable that is 1 when face m is assigned to cluster k and 0 otherwise. As an alternative to representing each cluster by its centroid, each cluster can be described as a Gaussian, but for our data, the resolved labels were not significantly different. The key is not necessarily how we represent each distribution, but how each face is assigned to a cluster.

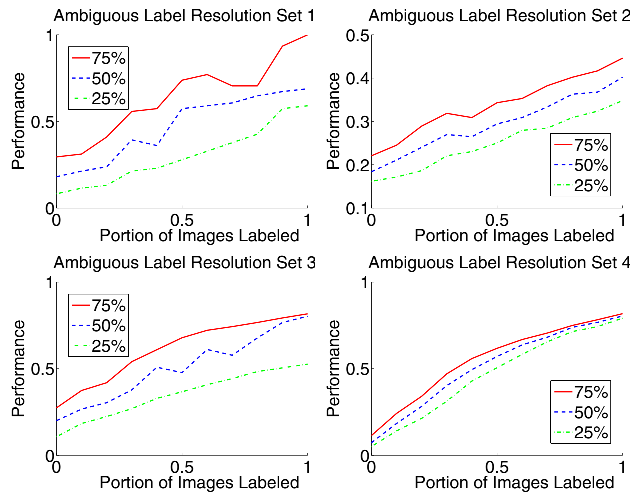

Figure 6. Performance of Ambiguous Label Resolution. The graphs show the median, 25% and 75% performances from 150 trials on each of four image collections as a function of the portion of the image collection that was ambiguously labeled.

4.1. Evaluation

The ambigous label resolution algorithm was applied to four consumer image collections. A portion of the images are randomly selected to be ambiguously labeled. Figure 12 shows an example of the resolved ambiguous labels for a set of images. The performance of the algorithm is quantified by finding the fraction of the number of all faces that are assigned the correct labels, and the results of a set of 150 trials with random initialization are shown in Figure 6. As expected, the performance of the algorithm improves as the number of ambiguously labeled images increases. It should be noted that the ALR algorithm, like k-means, is sensitive to the initial starting condition. In practice, multiple restarts are used and the start which converges to the minimum objective function is returned [5]. With our data ALR always converged, generally in fewer than 20 iterations.

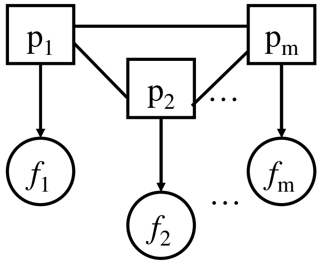

Figure 7. A graphical model that represents the features f and the people p in an image. Each person p_m has an undirected connection to all other people.

5. Classifying with Resolved Labels

Using the resolved labels, a classifier is trained for recognizing the individuals in the image collection. Of course, the resolved ambiguous labels contain some errors, so we must determine whether an effective classifier can be designed in the face of these erroneously labeled samples. We make the assumption that the people in the unlabeled images are in the set $\mathbf{N}$ of the unique individuals.

5.1. Images with one face

When an unlabeled image contains only a single face, the label for the face with features f is found according to Bayes rule:

$$p_{\text{MAP}} = \arg\max_{n \in \mathbf{N}} P(n|f) \tag{1}$$

$$= \arg\max_{n \in \mathbf{N}} P(f|n)P(n) \tag{2}$$

The distribution $P(f|n)$ is modeled with a Gaussian. When the computed covariance matrix is ill-conditioned, a generic covariance matrix, derived from many individuals, is substituted as the covariance matrix for that label. We have only ambiguously labeled images, so the Gaussians are computed using the resolved ambiguous labels.

The estimate of the prior probability $P(n)$ is derived from the ambiguous labels by counting the number of images containing a specific individual, according to:

$$P(n) = \frac{\sum_j y_{nj}}{\sum_u \sum_j y_{uj}} \tag{3}$$

where

$$y_{nj} = \begin{cases} 1 & n \in \Psi_j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

5.2. Images with multiple faces

The identities of multiple people in an image are not independent. There are two intuitive reasons for this. First, according to the unique object constraint, each individual can only appear once in the image. Second, multiple people in a consumer image generally have some kind of personal relationship. The group prior represents both the unique object constraint and the relationship between individuals that makes one group more likely to appear together in an image than another. For example, if we believe that Jen is in the image, then our belief that her brother Tommy is also in the image might increase. Thus, the classification of face identity should consider the features associated with all faces in the image.

Figure 7 graphically models the relationship between the identities of the people in the image and the observed features. The set of M people in the image is denoted $\mathbf{p}$, the set of all features is $\mathbf{f}$, and $\mathbf{n}$ is a subset of $\mathbf{N}$ with M elements and is a particular assignment of a name to each person in $\mathbf{p}$. A particular person in the image is p_m, the associated features are f_m, and the name assigned to person p_m is n_m. The joint probability $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$ of all the M people in a particular image, given the set of features is written:

$$P(\mathbf{p} = \mathbf{n}|\mathbf{f}) = \frac{P(\mathbf{f}|\mathbf{p} = \mathbf{n})P(\mathbf{p} = \mathbf{n})}{P(\mathbf{f})} \tag{5}$$

$$\propto P(\mathbf{p} = \mathbf{n}) \prod_m P(f_m|p_m = n_m) \tag{6}$$

Consistent with the model, we proceed from (5) to (6) by recognizing that the appearance of a particular person f_m is independent of all other individuals in the image once the identity of the individual p_m is known to be n_m. Tommy looks like Tommy regardless of who else is in the image.

Because we have access to a set of ambiguously labeled images, we can estimate the group prior $P(\mathbf{p} = \mathbf{n})$ or equivalently $P(\mathbf{n})$, the prior probability that a particular set $\mathbf{n}$ of M individuals would appear together in an image. First, we consider the case of estimating the group prior for any combination of two individuals:

$$P(n^u, n^v) = \frac{\sum_j y_{n^u j}\, y_{n^v j} + \alpha(u, v)}{\sum_{g,h \in N} \sum_j y_{n^g j}\, y_{n^h j} + \alpha(g, h)} \tag{7}$$

where

$$\alpha(u, v) = \begin{cases} \beta & u \neq v \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The function $\alpha(u, v)$ with a small non-zero β ensures that any two people have a non-zero probability of appearing together and at the same time respects the unique object constraint. The prior is estimated by counting the number of images that the pair n_u and n_v appear in together, divided by the total number of pairs of people in all images. One beautiful aspect is that this estimate is independent of the outcome of the ambiguous label resolution algorithm, so

$P(n^u, n^v)$ is the maximum likelihood estimate of the group prior. The size of $P(\mathbf{n})$ grows exponentially with the number of elements in $\mathbf{n}$, yet Figure 2 shows that images with increasing numbers of people are more rare. Instead of attempting to learn $P(\mathbf{n})$ for large M (i.e. $M > 2$) from the data, we estimate it from $P(n^u, n^v)$:

$$P(\mathbf{n}) = \frac{\prod_{u,v \in \mathbf{n}} P(n^u, n^v)}{\sum_{\mathbf{q} \subseteq \mathbf{N}} \prod_{u,v \in \mathbf{q}} P(n^u, n^v)} \tag{9}$$

where $\mathbf{q}$ has M elements. Equation (9) represents the group prior for any number of particular people appearing together in an image as a fully connected pairwise Markov model, again consistent with the model of Figure 7.

For a particular image with M people in an image collection of K unique individuals, there can be $\text{Vals}(\mathbf{n})$ different assignments of names to the people in the image.

$$\text{Vals}(\mathbf{n}) = \binom{K}{M} M! \tag{10}$$

$\text{Vals}(\mathbf{N})$ grows exponentially with both K and M, so we are relieved that both tend to be small so we can explicitly solve for $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$. For example, when $K = 7$ and $M = 5$, $\text{Vals}(\mathbf{N}) = 2520$.

Once $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$ is found, there are many different inference questions that can be answered by marginalizing the joint distribution.

5.2.1 Most Probable Explanation (MPE)

In MPE, the goal is to find the most probable labeling of all faces in the image. This assignment corresponds to the mode of $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$:

$$\mathbf{p}_{\text{MPE}} = \arg\max_{\mathbf{n} \in \mathbf{N}} P(\mathbf{p} = \mathbf{n}|\mathbf{p}) \tag{11}$$

5.2.2 Maximum Apriori Probability (MAP)

In MAP, the goal is to find the most probable identity of the m^{th} particular individual in the image. Therefore we marginalize over the name assignments of the other $M-1$ people in the image.

$$p_{m\text{MAP}} = \arg\max_{n_m^k \in \mathbf{N}} \sum_{p_i, i \neq m} P(\mathbf{p} = \mathbf{n}|\mathbf{f}) \tag{12}$$

5.2.3 Ambiguously Labeling

Inference can provide ambiguous labels for an unlabeled image. We desire to name the individuals in the image, but we do not specify which face is associated with which name. This would be particularly useful for auto-captioning the image.

$$p_{\text{AMB}} = \arg\max_{\mathbf{n} \in \mathbf{N}} \sum_{\mathcal{P}(\mathbf{n})} P(\mathbf{p} = \mathbf{n}|\mathbf{f}) \tag{13}$$

where $\mathcal{P}(\mathbf{n})$ denotes all permutations of the set $\mathbf{n}$.

5.2.4 Retrieval Based on Identity

Perhaps the most important query that could be posed is: Given the observed features f, what is the probability that a particular person n^q is in this image? This query has obvious applications for image retrieval based on whom the image contains. A query for images of a particular person can return images ranked according to $P(n^q|\mathbf{f})$. To satisfy this query, we simply sum $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$ over all sets of $\mathbf{n}$ where one p_m is assigned to n_m^q.

$$P(n^q|\mathbf{f}) = \sum_{\mathbf{n}, n^q \subset \mathbf{n}} P(\mathbf{p} = \mathbf{n}|\mathbf{f}) \tag{14}$$

5.3. Evaluation

Classification with the group prior was applied to the four image collections. Facial geometry features were extracted as described. One image is selected as the test image. A portion of the remaining images are ambiguously labeled and input to ALR. The group prior $P(\mathbf{n})$ is estimated from the ambiguous labels and each individual's feature distribution is represented by a Gaussian, using the resolved labels. For the test image, the joint probability $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$ is estimated using the features $\mathbf{f}$. Inference is performed on the test image to determine an MPE assignment for all faces in the image, a MAP assignment for each face, an assignment of ambiguous labels, and the probability that each individual from that image collection is present in the image. It should be stressed that in this evaluation, each ambiguously labeled image contains at least two people, so the entire system works without a single face ever being positively identified by a user. The goal is to show classification is improved with the group prior.

The results are shown in Figures 8 - 11. Figure 8 shows the results for MPE, where the performance is the percentage of test images that all faces were correctly identified as a function of the amount of ambiguously labeled data. Set 2 proves to be the most difficult because individuals from this image collection have a large amount of overlap in the feature space. Figure 9 shows the results for MAP, where the classification rate is the percentage of faces that were correctly classified. Figure 10 shows the results for ambiguously labeling the test images, where the classification rate is the number of images that are assigned the correct ambiguous labels. Four different priors were used in each experiment. The group prior is the full model that includes both the unique object constraint and the prior for specific groups of individuals. The UOC prior enforces the unique object constraint, but assumes that each group of individuals has equal probability of appearing in an image (we use $P_{\text{UOC}}(n^u, n^v) \propto \alpha(u, v)$, from (8)). The individual prior

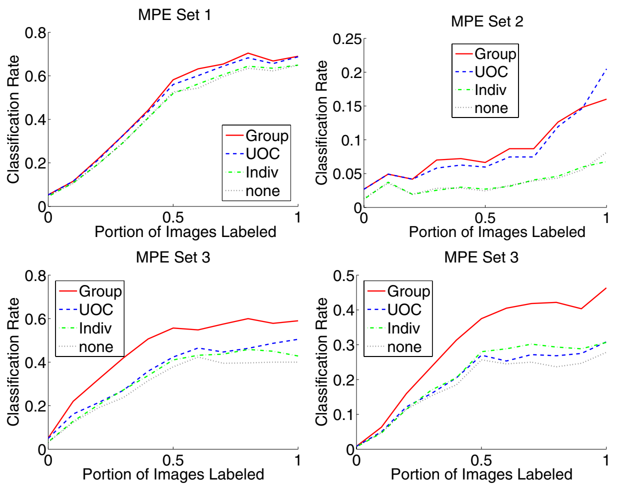

Figure 8. MPE performance on four consumer image collections using four different priors, as a function of the portion of the image collection with ambiguous labels.

(“Indiv”) considers only the prior probability of an individual appearing in an image, and finally no prior at all is used (“none”). When using the individual prior or no prior, each face is classified as if it were the only face in an image, according to (2). Inference using the group prior and the UOC prior considers the features of all faces in an image for inference. By representing the social relationships between individuals with the group prior, the performance is nearly always improved over the UOC prior, sometimes by as much as 10-15%.

Figure 11 shows the accuracy of using the system to produce the score $P(n^q|\mathbf{f})$ that would be useful for an image retrieval system. The performance using the resolved ambiguous labels is compared against using the actual ground truth labels, which is the upper bound for the performance of the ALR algorithm. The score $P(n^q|\mathbf{f})$ is produced for each test image for each individual in the set $\mathbf{N}$. Precision-recall curves are generated by varying a threshold on $P(n^q|\mathbf{f})$. All images except the randomly selected test image are ambiguously labeled. Mistakes made in resolving the ambiguous labels hurt the performance, but the recognition rates are surprisingly good, again considering that not a single face was explicitly labeled with the correct name.

6. Discussion

We have introduced the problem of ambiguously labeled images in the context of labeling people in consumer image collections. We described an algorithm for resolving the ambiguous labels. Using the ambiguous labels, we learn a group prior for classification of people in unlabeled images. The group prior enforces the unique object constraint that an individual can appear at most one time in an image and

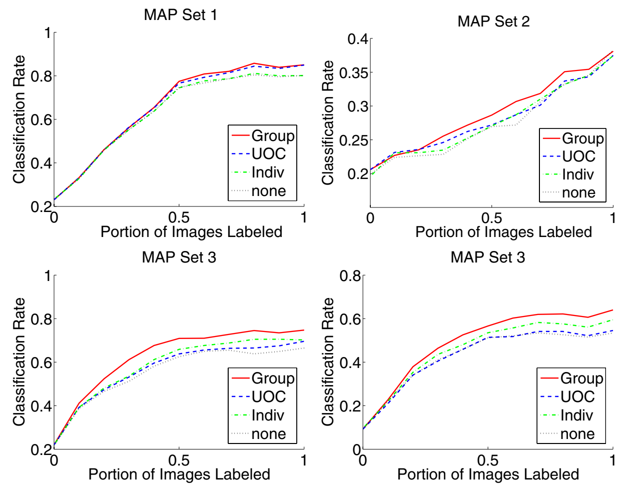

Figure 9. MAP performance on the four image collections.

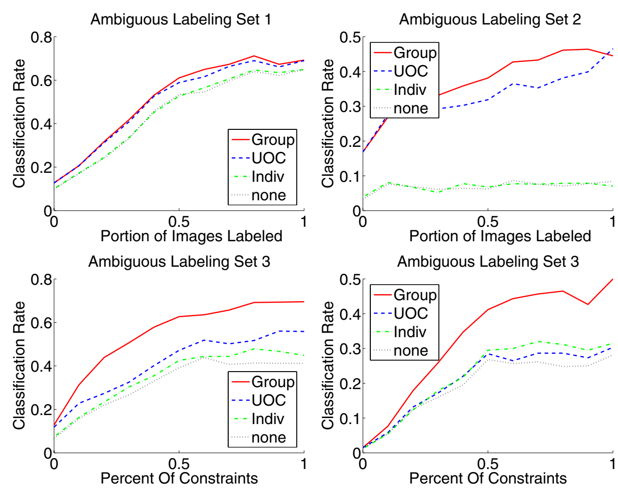

Figure 10. Performance of ambiguous labeling on the four image collections.

indicates the probability of specific groups of people appearing together in an image. We demonstrated that despite errors in resolving ambiguous labels, useful classifiers can be trained with the resolved labels. By modeling the relationships between people in an image with the group prior, classification performance is significantly improved in all of our test sets.

7. Future Work

We are expanding on this work, which models the relationship between people in a single image, to a model which includes all faces in the image collection. Each face from the image collection is a node in a large undirected network, and links connect other similar faces (based on distance in feature space) and dissimilar faces (other faces

Figure 12. An example of automatically resolved ambiguous labels for 15 images. Only two images contain mistakes, the last image of the first row, and the fourth image in the second row.

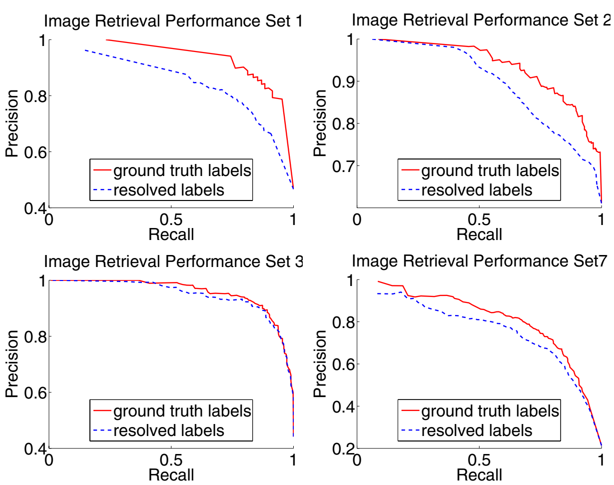

Figure 11. Retrieval performance on the four consumer image collections. In both cases the group prior is used. Training with resolved labels, which contain some mistakes, hurts the performance but the results are still very good.

from the same image). We expect approximate inference will allow the use of group prior even when the number of people is large. Furthermore, we plan to expand and release our database of labeled image collections.

References

[1] T. Berg, A. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. Forsyth. Names and faces in the news. In *Proc. CVPR*, 2004. 1, 2

[2] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *CVIU*, 1995. 3

[3] M. Jones and P. Viola. Fast multiview face detector. In *Proc. CVPR*, 2003. 3

[4] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society of Industrial and Applied Mathematics*, 1957. 4

[5] M. Murty, A. Jain, and P. Flynn. Data clustering: A review. *ACM Comput. Surv.*, 1999. 4

[6] M. Naaman, R. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *Proc. JCDL*, 2005. 2

[7] A. Ng, Y. Weiss, and M. Jordan. On spectral clustering: Analysis and an algorithm. In *Proc. NIPS*, 2002. 2

[8] S. Satoh and T. Kanade. Name-it: Association of face and name in video. In *Proc. CVPR*, 1997. 2

[9] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 2000. 2

[10] A. Torralba and P. Sinha. Statistical context priming for object detection. In *Proc. ICCV*, 2001. 2

[11] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl. Constrained k-means clustering with background knowledge. In *Proc. ICML*, 2001. 2

[12] S. Yu and J. Shi. Grouping with bias. Technical report, CMU, 2001. 2

[13] L. Zhang, Y. Hu, M. Li, and H. Zhang. Efficient propagation for face annotation in family albums. In *Proc. MM*, 2004. 2

[14] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 2003. 2, 3