Family Member Identification from Photo Collections

Qieyun Dai[1,2] Peter Carr[2] Leonid Sigal[2] Derek Hoiem[1]
[1]University of Illinois at Urbana-Champaign [2]Disney Research
dai9@illinois.edu {peter.carr, lsigal}@disneyresearch.com dhoiem@uiuc.edu

Abstract

Family photo collections often contain richer semantics than arbitrary images of people because families contain a handful of specific individuals who can be associated with certain social roles (e.g. father, mother, or child). As a result, family photo collections have unique challenges and opportunities for face recognition compared to random groups of photos containing people. We address the problem of unsupervised family member discovery: given a collection of family photos, we infer the size of the family, as well as the visual appearance and social role of each family member. As a result, we are able to recognize the same individual across many different photos. We propose an unsupervised EM-style joint inference algorithm with a probabilistic CRF that models identity and role assignments for all detected faces, along with associated pairwise relationships between them. Our experiments illustrate how joint inference of both identity and role (across all photos simultaneously) outperforms independent estimates of each. Joint inference also improves the ability to recognize the same individual across many different photos.

1. Introduction

Digital cameras make it easy to acquire large photo collections, creating a need for good tools to organize them. While most existing approaches focus on low-level information [23], such as measures of image quality and saliency, more semantic understanding of image collections is required for an ever growing diverse set of tasks (*e.g.*, summarization, visual story telling, visual search, *etc.*). A key component to this semantic understanding is the ability to recognize identities and roles (for a given scenario) within the photo collection. We address the problem of unsupervised family member discovery from an unlabeled collection of personal photographs. We discover family members by identifying faces across multiple photos having consistent appearance and social role (*e.g.*, child, father, mother); see Figure 1.

The problem of family member identification is related to face recognition and verification. One challenge is that we do not have supervised identity data. To address this problem, we cluster faces and use an EM-style approach to refine our face clusters. Typically, face verification algorithms consider the similarity of one face to another or to a trained model. Our scenario differs in that each photo collection includes a small number of frequently occurring faces, providing an opportunity for more robust identity assignment. However, family members tend to have similar faces, which can make it difficult to distinguish individuals in widely varying casual photographs. We reason about all faces within a photo collection simultaneously, which allows us to incorporate mutual exclusion constraints. For instance, each identity cannot appear in each photo more than once. Another difference with respect to traditional face recognition and verification is that we introduce the concept of social roles into our pipeline, similar to [19]. However, unlike [19], where a weakly supervised scenario is assumed, we assume no supervision. We show that the introduction of role information helps improve face verification results compared tos appearance-based approaches (*e.g.*, [12]).

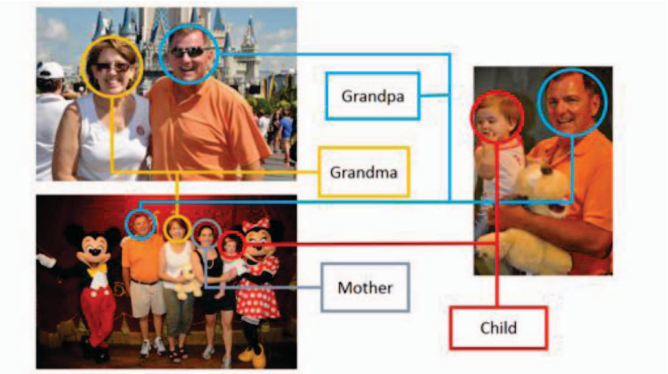

Figure 1. **Family member identity and role discovery.** Our goal is to find the same face across all images within a photo collection while simultaneously inferring the social role of each person.

Family member discovery is important because it provides an alternative way of looking at face recognition/verification in family photo collections. Rather than labeling the identify of each face (which requires active user input), role provides an unsupervised way of understand-

978-1-4799-6683-7/15 $31.00 © 2015 IEEE
DOI 10.1109/WACV.2015.136

IEEE computer society

ing each person and the composition of the family, which in itself may be more useful (*e.g.*, for e-commerce). Furthermore, family composition and family member identification is also one of the key steps towards organization and browsing family photographs. Knowing roles and identities would allow users to browse collections based on who they want to see or allow diverse preview of a collection by choosing photos containing different family members. In addition, knowledge of *actors* is essential in generation of semantic storylines [8, 17, 22] from photo collections.

Contributions: Unlike traditional face recognition or verification approaches that look at each face or pair of faces individually, we propose a framework that reasons about all faces in a photo collection simultaneously. This insight allows us to re-identify the same person more reliably than traditional algorithms based only on face similarity. We show that joint modeling and inference over role and identity is mutually beneficial: role knowledge improves identity clustering quality, and identity information helps improve role assignment accuracy. We also show that incorporating clothing appearance improves the overall accuracy of face verification.

1.1. Related Work

Recently, several works have used a social perspective when analyzing photos or videos of groups of people. Lee *et al*. [11] leverage "social context" of co-occurring people to discover novel faces in untagged photos. They show that given an unknown face, by looking at its co-occurrence with known faces, the performance of their system increased greatly. Lin *et al*. [13] present a framework that jointly tags people, events, and locations in photos using a generic probabilistic context model that links different domains through a set of cross-domain relations. Murilloy *et al*. [16] build a graph of groups of people to learn models of urban tribes (substructures of people who share common interests and tend to have similar styles of clothes and behavior).

Xia *et al*. [21] study the problem of "child-parent" verification in a photo using a transfer subspace learning approach. Ding and Yilmaz [3] infer social relations among actors in a video (*i.e*., grouping people into different communities) using visual concepts such as "shooting", "ship", and "beach". Ramanathan *et al*. [18] design a CRF model to encode inter-role interaction and person-specific social descriptors to recognize social roles played by people in a video in a weakly supervised fashion.

Most closely related is the work of Wang *et al*. [19], which uses face detection features (such as relative image position and age difference) for social relationship classification and incorporates social relationships into face recognition. Unlike from our approach, they assume a known family size and access to weakly labeled photos during the training stage, which makes the identity classification problem easier. In addition, they reason about each image in isolation, while we consider all faces in all photographs simultaneously. Our holistic approach allows us to make use of consistency constraints such as the same person cannot appear more than once in an image, and all faces assigned the same identity should also be assigned the same role.

We build on ideas from Gallagher and Chen [5] who use clothes co-segmentation to help recognize people. The paper argues that facial similarity is often insufficient to tell two people apart, which is especially true for family photo collections, since family members are related to each other and often share certain facial traits. We also incorporate clothing appearance information to improve face verification performance.

Berg *et al*. [1] also examine photo collections and perform face clustering on a large dataset of captioned news images. Different from our approach, they do not consider the social relationship between faces, and make use of the names extracted from captions.

2. Approach

We define a family as a collection of F individuals, where each individual fulfills one of five social roles $\mathcal{R} = \{\mathsf{child}, \mathsf{father}, \mathsf{mother}, \mathsf{grandfather}, \mathsf{grandmother}\}$. A family photo collection is a set of M photos containing a total of N face detections $\mathcal{D} = \{D_1, \ldots, D_N\}$ distributed among the images. Each detected face D_i has an unknown role R_i and identity I_i which we must estimate. To simplify the problem, we assume a family can have multiple children but at most one father, one mother, one grandfather and one grandmother[1]. Furthermore, we define a set of K generic identity labels $\mathcal{I} = \{I_1, \ldots, I_K\}$ by grouping the detected faces $\mathcal{D}$ into visually similar clusters. Typically, the number of automatically discovered clusters is an over segmentation of identity—*i.e*. $K > F$. Therefore, in addition to estimating the identity and role of each face detection, we must also estimate the number of family members by iteratively merging face clusters so that $K \approx F$.

We model the relationships between all face detections using a conditional random field with following potentials:

- E_{id}: Unary potential indicating how likely a face belongs to a identity/cluster. (Section 2.1)
- E_{role}: Unary potential indicating how likely a face has a certain social role. (Section 2.2)
- $E_{similarity}$: Binary potential measuring the similarity between two faces from different images. (Section 2.4)
- E_{unique}: Binary potential ensuring each identity appears at most once in each image.

[1]Note that these assumptions can easily be relaxed and our overall approach is not specific to them or the chosen role labels.

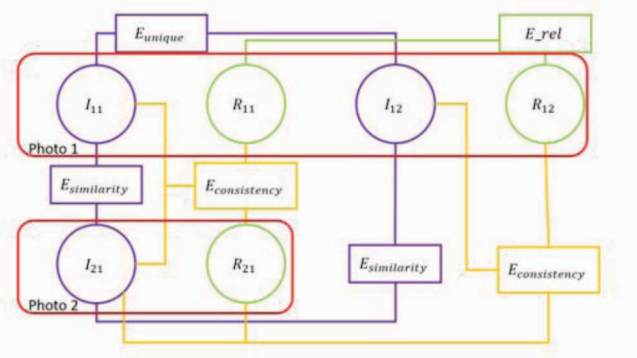

Figure 2. **Factor Graph**. A factor graph over two images with three face detections illustrating our objective function.

- $E_{relationship}$: Binary potential measuring how likely a pair of faces takes on a pair of role labels. It also ensures each role (except child) only appears once in an image. (Section 2.3)

- $E_{consistency}$: Quaternary potential to ensure that if two faces are assigned the same identity/cluster, they are also assigned the same role.

Our goal is to find the joint assignment $(\mathbf{I}^*, \mathbf{R}^*)$ of identities and roles that minimizes the objective energy function

$$\begin{aligned} E(\mathbf{I},\mathbf{R}) = &\sum_{i=1}^{N} E_{id}(I_i) + \sum_{i=1}^{N} E_{role}(R_i) \\ &+ \sum_{\substack{i,j \text{ in} \\ \text{same image}}}^{N} \Big[E_{unq}(I_i, I_j) + E_{rel}(R_i, R_j) \Big] \\ &+ \sum_{\substack{i,j \text{ in} \\ \text{diff. images}}}^{N} \Big[E_{sim}(I_i, I_j) + E_{con}(I_i, I_j, R_i, R_j) \Big]. \end{aligned} \tag{1}$$

A factor graph depicting the case of two images with three face detections is shown in Figure 2.

2.1. Identity Clustering

Since the number of family members is not known, clustering methods such as k-means are infeasible. Instead, we employ an iterative approach borrowing ideas from exemplar SVMs [14]. We begin with all detected faces $\mathcal{D}$ unclustered. We select one face as a positive example from the center of the largest cluster generated by affinity propagation [4]. We then train an SVM using this positive example and an initial negative set of faces randomly selected from Labeled Faces in the Wild [7]. The SVM is then used to classify all faces in the family photo collection. We enlarge the training set by adding all faces from our photo collection with a predicted probability ≥ 0.8 to the positives, and all faces with a predicted probability ≤ 0.2 to the negatives. We repeatedly retrain the SVM until there are no changes to the training set. We then pick a new exemplar from the unassigned (or negative) faces in our set and repeat the above process until the size of the negative set drops below 3.

Then we apply a pruning step to the generated clusters. We merge clusters whose members overlap more than 70% (all examples with predicted probability ≥ 0.5 are considered a member of the cluster) and discard clusters with fewer than 3 members. The remaining K clusters represent the initial prediction of identities $\mathcal{I}$. In our implementation, we use the probabilistic version of libsvm [2] and use FPLBP features [20] extracted on aligned faces [6].

2.2. Role Classification

We use the gender, age, and age uncertainty values produced by the face detector [10] to train five 1-versus-all linear SVM classifiers (one for each role). The training data is divided into a characterization set (for generating histograms of feature values) and a training set (for training the SVM classifier).

We use the characterization set to estimate $P(\text{age}|R_i)$, $P(\text{ageDev}|R_i)$ and $P(\text{gender}|R_i)$ for each role using histogram frequency counts. For age, we use 12 bins with a bin width of 5, and all predicted ages larger than 60 are clamped at 60. For age deviation, we use 10 bins with a bin width of 2, and limit all age deviations to 20. For gender prediction, we use two bins, male and female.

For each example in the training set we compute its probability of belong to each role given the predicted age, age deviation, and gender. By assuming an equal prior for each role, the feature vector is:

$$\begin{aligned} \Big[&P(\text{age}|\mathsf{child}), P(\text{age}|\mathsf{father}), \ldots \\ &P(\text{ageDev}|\mathsf{child}), P(\text{ageDev}|\mathsf{father}), \ldots \\ &P(\text{gender}|\mathsf{child}), P(\text{gender}|\mathsf{father}), \ldots \Big]. \end{aligned}$$

2.3. Relationship Classification

There are 25 possible ordered pairwise role relationships. However, since each role (except child) can only appear once in the same image, relationship pairs such as "father-father" cannot exist. Furthermore, we combine the "father-mother" and "grandfather-grandmother" relationships into a "husband-wife" pair and an equivalent "wife-husband" pair for "mother-father" and "grandmother-grandfather". We follow [19] and extract the following features:

- height difference between the two faces
- distance between the two faces

- size ratio between the two faces
- age difference
- gender prediction

Both height difference and distance between faces are measured relative to the average face size in the given image. Similarly, we train a one-vs-all linear SVM for each relationship using the above features.

2.4. Face Recognition

We gauge whether a pair of faces from different images correspond to the same individual by training an SVM utilizing both appearance and role information.

For appearance information, we consider both facial similarity and clothing similiarity, since members of the same family often share certain facial traits but are generally dressed differently. We estimate face similarity using the chi-square distance between FPLBP features [20] extracted on aligned faces [6] and clothing similarity using the intersection of L*a*b histograms computed on segmented clothes regions generated by graph cuts (please refer to the supplemental materials for details on clothes segmentation).

In addition to clothing and facial similarity scores, we use the verification score provided by Li *et al*. [12] as an additional feature, where they proposed a probabilistic elastic matching algorithm with an additional joint Bayesian adaption component to estimate whether two faces correspond to the same individual. To utilize role information, we represent the role prediction for each face as a five element vector $[P_{\mathsf{child}}, P_{\mathsf{father}}, P_{\mathsf{mother}}, P_{\mathsf{grandfather}}, P_{\mathsf{grandmother}}]$ where each element is the probability output of the corresponding role classifier. The role prediction distance between two roles is calculated as the Euclidean distance between two role prediction vectors.

Our supplemental material describes how use of clothing and role information perform better than using only the verification score [12] for classifying pairs of faces of family members.

2.5. Inference

Inference over clusters and labels is difficult because high-order constraints on the labeling depend on the clustering. We use a two step approach to obtain an approximate solution to the inference problem using the TRW-S algorithm [9]. We begin by estimating the identities of all detected faces, and then infer role assignments afterwards.

Identity Inference. Here, we minimize the terms of Equation (1) which only directly depend on identity

$$\sum_{i=1}^{N} E_{id}(I_i) + \sum_{\substack{i,j \text{ in} \\ \text{same image}}}^{N} E_{unq}(I_i, I_j) + \sum_{\substack{i,j \text{ in} \\ \text{diff. images}}}^{N} E_{sim}(I_i, I_j). \quad (2)$$

In practice, the hard constraints in $E_{unique}(\cdot)$ are encoded as large positive energies (in our implementation, 1e10) whenever two faces in the same image are assigned the same identity.

Role Inference. Since faces in the same identity cluster are considered the same person, they should be assigned the same role. Therefore, we create K new variables (K is the number of identity clustered generated in step 1) representing the predominant role label assigned to each identity cluster. We define a new binary potential $E_{cluster}(R_i, R_k)$ to penalize cases when the role assigned to a face differs from the role assigned to its cluster

$$E_{cluster}(R_i, R_k) = \begin{cases} \alpha & \text{if } R_i \neq R_k, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Here, R_i is the role assigned to the i^{th} face, and R_k is the predominant role assigned to the k^{th} cluster. Since the identity clustering may not have perfect purity, α is generally smaller than $E_{unique}(I_i = I_j)$. In our implementation, we set α to 100.

To estimate role assignments, we minimize the following terms in Equation (1) where $E_{con}(I_i, I_j, R_i, R_j)$ is approximated by $E_{cluster}(R_i, R_k)$ and the current identity assignments are fixed:

$$\sum_{i=1}^{N} E_{role}(R_i) + \sum_{\substack{i,j \text{ in} \\ \text{same image}}}^{N} E_{rel}(R_i, R_j) + \sum_{\substack{i \\ k=I_i}}^{N} E_{cls}(R_i, R_k). \quad (4)$$

2.6. Iterative Update

After performing the two-step inference, we re-cluster the faces in an attempt to converge on the correct family size—*i.e*. $K \approx F$. We start with the identity assignments generated by the inference algorithm, and iterate through two steps: 1) building identity models 2) face assignment.

To build identity models, we train one linear SVM per identity cluster using all faces having that identity as positive training data and the rest of the faces as negative training data. Then we test each identity classifier on all the faces to get the probability a face belongs to this cluster. The faces are then re-assigned to the cluster with the highest probability. This step is repeated until the assignment of faces to clusters do not change. The features used for retraining the identity models are the same as in Section 2.1.

We generate a new set of K identity clusters based on the final assignment, and repeat the inference with these new identities.

2.7. Post Processing

Since the second stage of our inference algorithm does not require each identity cluster to have a unique role label,

Set	# Id	# Child	# Imgs	# Faces	Roles
1	4	1	60	107	C, F, M, GM
2	2	0	70	95	F, M
3	4	2	101	192	C, F, M
4	6	4	58	119	C, F, M
5	5	1	84	139	C, F, M, GM, GF
6	4	1	120	214	C, M, GP, GM
7	2	0	58	101	F, M

Table 1. **Our dataset:** Statistics for the 7 photo collections used for testing. Each column shows: photo collection number, family size, number of children, total number of images and faces in this collection, and the list of social roles in this family. *Roles* shows which roles exist in this particular collection, where C = Child, F = Father, M = Mother, GM = Grandmother and GF = Grandfather.

we apply a final post-process merging all non-child identity clusters having the same role. Furthermore, in the rare case where the faces constituting an identity cluster have not been assigned consistent role labels, we split that identity cluster based the role assignment.

3. Experiments

We first give an overview of the datasets used in our experiments (Section 3.1), then we evaluate our approach from three aspects: (1) how role information helps identity prediction (Section 3.2), (2) how identity information helps role prediction (Section 3.3), (3) evaluation of the entire framework that performs joint inference (Section 3.4).

Inspired by the experimental design of [19], we only keep detected faces that correspond to ground truth family members. Unlike [19], however, we do not manually add missed faces and do not provide weak supervision in the form of name lists for the images. Our identity clustering method is stochastic and sometimes returns slightly different cluster structures. Therefore, we report average values over five different runs of all experiments.

3.1. Dataset

We make use of the Gallagher Collection Person Dataset [5] as part of our training data for role and relationship classifiers. We use the LFW [6, 7] as our initial negative set for training identity classifiers.

In addition, we create a dataset of our own containing photo collections of 16 different families taken at amusement parks. These collections cover families of different composition and size. We use 9 of the 16 sets for training, and the remaining 7 for testing (see Table 1 for summary).

Each person in our dataset is annotated with an identity label, a bounding box indicating location of the face and the body skeleton. For family members, the identity annotation also shows their social role, for example: "child1", "father", *etc*., while non family members are given identities such as "femaleAdult1", "maleAdult1". Figure 3 shows example annotations on one image.

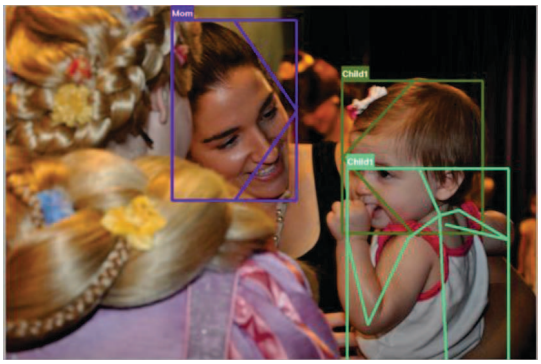

Figure 3. **Dataset annotations:** available on our dataset. Each person is annotated with a role/identity label, a bounding box around his face, and a skeleton for his body.

3.2. Identity Clustering

To evaluate identity clustering, we follow [15] and measure *purity* and *normalized mutual information* (NMI) for the generated clusters (Table 2).

Error Metrics: Purity is computed as

$$Purity(C, G) = \frac{1}{N} \sum_i max_j |c_i \cap g_j|$$

where C represents the set of predicted clusters and G is the set of ground truth clusters, N is the total number of examples, and c_i represents the i-th predicted cluster and g_j the j-th ground truth cluster.

Since making each example a cluster of its own would yield perfect purity, we use normalized mutual information (NMI) to trade off the quality of clustering against the number of clusters.

$$NMI(C, G) = \frac{I(C, G)}{(H(C) + H(G))/2}$$

I(C, G) is the mutual information between C and G,

$$I(C, G) = \sum_i \sum_j P(c_i \cap g_j) \log \frac{P(c_i \cap g_j)}{P(c_i)P(g_j)},$$

where $P(c_i)$ is the probability that a face ends up in the i-th predicted cluster, $P(g_j)$ is the probability that an face has the j-th ground truth label, and $P(c_i \cap g_j)$ is the probability that an example is in the i-th predicted cluster and has the j-th ground truth label. I(C, G) is 0 when the generated clusters are random with respect to the ground truth, and reaches its maximum when the generated clusters have perfect purity.

H is entropy defined as

$$H(S) = -\sum_i P(s_i) \log P(s_i)$$

which penalizes generating too many clusters.

Baselines: To test how, and if, role prediction helps improve identity prediction, we compare results produced by the proposed method against outputs of (1) the clustering algorithm described in Section 2.1 and (2) an iterative update baseline based solely on identity information. The second baseline is a simplified variant of our approach that does not take role information into account. As such, we run only stage 1 of the inference framework defined in Sec 2.5 and instead of using face similarity score generated by our classifier (Section 2.4) (which uses role prediction as one feature), it uses the score produced by [12]. We show in the supplemental material that by incorporating role information, our pairwise face similarity classifier achieves an average accuracy of 0.729 and an average area under the ROC curve of 0.810, outperforming [12]'s 0.690 and 0.732 respectively.

Results and analysis: The average cluster purity and NMI score for each photo collection is shown in Table 2. We can see that, for almost cases, the proposed approach outperforms the two baselines by a large margin. Note that the results for Identity Classifier differs from the results in Table 4 since there is no post-processing step. Also note the reduction in the number of clusters achieved by our method. In many cases we can achieve better performance while having 2 to 3 times fewer clusters than our strong baseline that does not take role predictions into account.

3.3. Role Classification

Error Metrics: Average accuracy of role classification can be misleading because some roles are much more common than others. Therefore, we measure precision and recall for each role across all photo collections (Table 3) and show role confusion matrix in Figure 4.

Baselines: Similar to Section 3.2, we compare against output of the role classifier in Section 2.2 and a baseline using only role and relationship information which tries to minimize:

$$\sum_{i=1}^{n} E_{role}(R_i) \; + \sum_{\substack{i,j \text{ in} \\ \text{same image}}} E_{rel}(R_i, R_j).$$

Results and analysis: By looking at the confusion matrix (Figure 4), it's clear that the major source of error for both the "role classifier" and the "role only inference model", is the confusion between adult roles of the same gender. This is because both our role and relationship classifier depend on age prediction, but our face detector is less accurate at age prediction for grownups. This confusion is greatly reduced by our joint iterative approach, especially for Grandfather, where the recall increased by a large margin.

3.4. Inference Results

In this section, we evaluate the performance of the proposed joint iterative approach by comparing against two baseline approaches.

No Inference: In this baseline, we only perform identity clustering and role classification, treating each face independently. Each face is assigned the most likely id and role based on classifier outputs, and the assignment is post-processed as described in Section 2.7.

Single Round Inference: Instead of iteratively updating our identity models and re-running the inference, only one round of inference is performed. The same post-processing (Section 2.7) is applied to output of the baseline inference algorithm.

Table 4 shows role accuracy, identity cluster purity and identity cluster normalized mutual information (NMI) for all 7 test photo collections. Our role accuracy is better (in some cases by as much as 19%) than the baselines except for Set 4. The identity assignment is also generally better.

3.5. Qualitative Results

In this section, we show results produced on Sets 6 and 7 where Set 6 is a family of 5 with both grandparents and a young child while Set 7 is a young couple. Note that we use the general role labels "father" and "mother" for young couples, regardless of whether children are present in the photo collection or not.

We show qualitative results from three different perspectives. Figure 5 shows predictions on the same image before and after the inference step. Figure 6 shows a few successful examples where each face is assigned the correct role, and Figure 7 shows some typical failure examples. There are two main causes for error: (1) one face is assigned to the wrong cluster during identity prediction and thus given the wrong role, or (2) when one person is split into several predicted clusters, generating additional child clusters.

4. Summary

In this paper, we proposed an approach to jointly infer identities and social roles of faces in a family photo collection. We show that role information helps identity prediction and vice versa. Interesting directions for future work are (1) to add roles for non-family members and additional relatives, such as aunts and uncles, and (2) to consider other types of social groups, such as friends.

References

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

		Identity Classifier			Identity Only Inference			Joint Iterative Approach		
Set	# gt ids	Purity	NMI	# clusters	Purity	NMI	# clusters	Purity	NMI	# clusters
1	4	0.944	0.620	11	**0.953**	**0.841**	5	0.901	0.788	4
2	2	0.794	0.216	5	0.842	0.322	5	**0.983**	**0.853**	3
3	4	0.744	0.309	18	**0.789**	0.460	18	0.774	**0.500**	8.4
4	6	**0.595**	0.343	9	0.556	**0.345**	9	0.540	0.289	5.6
5	5	0.842	0.402	15	**0.927**	0.641	15	0.919	**0.745**	7.2
6	2	**0.865**	0.448	19	0.801	0.514	16	0.862	**0.636**	5.4
7	2	0.931	0.318	14	0.962	0.411	11.8	**0.984**	**0.778**	3.8
Average	-	0.816	0.380	-	0.833	0.505	-	**0.852**	**0.656**	-

Table 2. **Identity clustering average purity and NMI.** "# gt id" is the number of ground truth identities and "# clusters" is the number of identity clusters generated by the given method."Identity Classifier" refers to the method proposed in Sec 2.1, "Identity Only Inference" is the proposed baseline which does not use any role information. The results are averaged over five runs.

		Role Classifier		Role-Rel Inference		Joint Iterative Approach	
Role	Num Faces	Precision	Recall	Precision	Recall	Precision	Recall
Child	420	0.824	0.769	0.805	0.817	**0.947**	**0.903**
Father	155	0.539	0.723	0.587	0.587	**0.884**	**0.839**
Mother	224	0.519	0.478	0.551	0.580	**0.751**	**0.883**
GrandFather	42	0.198	0.381	0.333	0.524	**0.610**	**0.948**
GrandMother	126	0.363	0.230	0.405	0.270	**0.753**	**0.540**
Average	-	0.488	0.516	0.536	0.557	**0.789**	**0.822**

Table 3. **Evaluation of role assignment performance in terms of precision and recall.** "Role Classifier" is the method introduced in Section 2.2, "Role-Rel Inference" is single image based inference using only role and relationship information. Averaged over five runs.

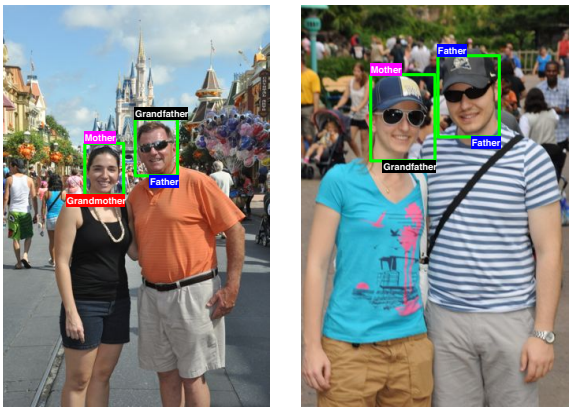

Figure 5. **Predictions before and after the inference step:** The text above the bounding box is the predicted role by our approach, and the text below the face is the role predicted by the role classifier. Different roles are shown in text boxes of different colors. The first image is taken from the set 6 and the second from set 7.

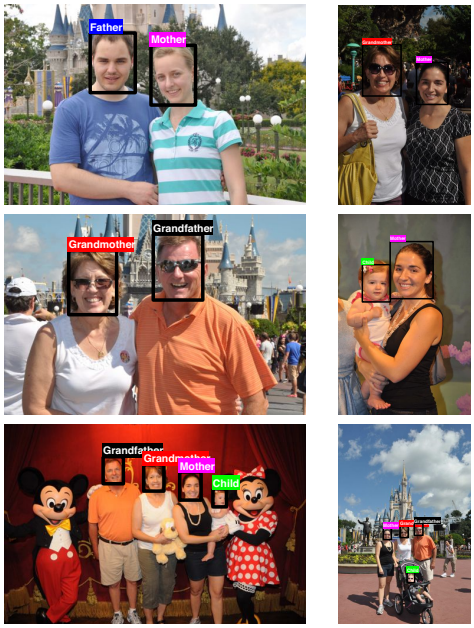

Figure 6. **Examples of performance:** Example images of successful role/identity labeling. The first image is from Set 7 and the remaining 5 images are from Set 6. Role labels are shown in text boxes of different above the face bounding box.

[2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011.

[3] L. Ding and A. Yilmaz. Inferring social relations from visual concepts. In *ICCV*, 2011.

[4] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 2007.

[5] A. Gallagher and T. Chen. Clothing cosegmentation for recognizing people. In *CVPR*, 2008.

[6] G. B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *ICCV*, 2007.

[7] G. B. Huang, M. Ramesh, and E. L.-M. T. Berg. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007.

[8] G. Kim and E. P. Xing. Jointly aligning and segmenting multiple web photo streams for the inference of collective photo storylines. In *CVPR*, 2013.

[9] V. Kolmogorov. Convergent tree-reweighted message pass-

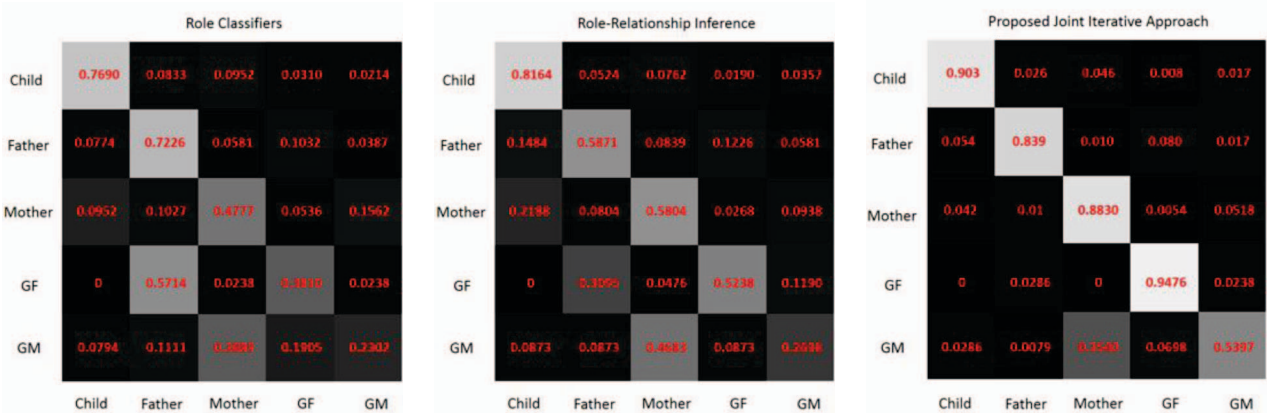

Figure 4. **Confusion matrices for role predictions:** The rows represent ground truth labels and columns predictions. Averaged 5 runs.

	No Inference			Single Round			Joint Iterative Approach		
Set	Role Acc	Id Purity	Id NMI	Role Acc	Id Purity	Id NMI	Role Acc	Id Purity	Id NMI
1	0.645	0.785	0.447	0.753	**0.927**	**0.814**	**0.781**	0.901	0.788
2	0.632	0.874	0.303	0.924	0.956	0.701	**0.960**	**0.983**	**0.853**
3	0.766	0.715	0.291	0.873	0.681	0.380	**0.923**	**0.774**	**0.500**
4	0.555	0.580	**0.332**	**0.714**	**0.543**	0.287	0.708	0.540	0.289
5	0.755	0.835	0.410	0.885	**0.927**	**0.745**	**0.907**	0.919	**0.745**
6	0.411	0.710	0.330	0.732	0.859	0.580	**0.742**	**0.862**	**0.636**
7	0.515	0.861	0.276	0.735	0.939	0.568	**0.929**	**0.984**	**0.778**
Avg	0.611	0.766	0.341	0.802	0.833	0.582	**0.850**	**0.852**	**0.656**

Table 4. **Results on role and identity assignment:** Left to right: results from identity and role classifiers after post processing, results after one round of inference and post-processing and resulting produced by the proposed iterative joint inference. Averaged over five runs.

Figure 7. **Failure Examples:** This figure shows cases where faces are assigned the wrong role label. The first image is from Set 6 and the last two image are from Set 7.

ing for energy minimization. *PAMI*, October 2006.

[10] C. Küblbeck, T. Ruf, and A. Ernst. A modular framework to detect and analyze faces for audience measurement systems. In *GI Jahrestagung*, 2009.

[11] Y. Lee and K. Grauman. Face discovery with social context. In *BMVC*, 2012.

[12] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In *CVPR*, 2013.

[13] D. Lin, A. Kapoor, G. Hua, and S.Baker. Joint people, event, and localization recognition in personal photo collections using cross-domain context. In *ECCV*, 2010.

[14] T. Malisiewicz, A. Gupta, and A. A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *ICCV*, 2011.

[15] C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*, chapter Flat clustering. Cambridge University Press, 2008.

[16] A. Murilloy, I. S. Kwakz, L. Bourdevx, D. Kriegmanz, and S. Belongie. Urban tribes: Analyzing group photos from a social perspective. In *CVPR*, 2012.

[17] P. Obrador, R. Oliveria, and N. Oliver. Supporting personal photo storytelling for social albums. In *ACM Multimedia*, 2010.

[18] V. Ramanathan, B. Yao, and L. Fei-Fei. Social role discovery in human events. In *CVPR*, 2013.

[19] G. Wang, A.Gallagher, and D. F. J. Luo. Seeing people in social context: Recognizing people and social relationships. In *ECCV*, 2010.

[20] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Post ECCV workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition*, 2008.

[21] S. Xia, M. Shao, J. Luo, and Y. Fu. Understanding kin relationships in a photo. *IEEE Transactions on Multimedia*, 2012.

[22] J. Yang, J. Luo, J. Yu, and T. Huang. Photo stream alignment and summarization for collaborative photo collection and sharing. *IEEE Transactions on Multimedia*, 2012.

[23] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung. Personalized photograph ranking and selection system. In *ACM Multimedia*, 2010.