Seeing People in Social Context: Recognizing People and Social Relationships

Gang Wang[1], Andrew Gallagher[2], Jiebo Luo[2], and David Forsyth[1]

[1] University of Illinois at Urbana-Champaign, Urbana, IL
[2] Kodak Research Laboratories, Rochester, NY

Abstract. The people in an image are generally not strangers, but instead often share social relationships such as husband-wife, siblings, grandparent-child, father-child, or mother-child. Further, the social relationship between a pair of people influences the relative position and appearance of the people in the image. This paper explores using familial social relationships as context for recognizing people and for recognizing the social relationships between pairs of people. We introduce a model for representing the interaction between social relationship, facial appearance, and identity. We show that the family relationship a pair of people share influences the relative pairwise features between them. The experiments on a set of personal collections show significant improvement in people recognition is achieved by modeling social relationships, even in a weak label setting that is attractive in practical applications. Furthermore, we show the social relationships are effectively recognized in images from a separate test image collection.

1 Introduction

Personal image collections now often contain thousands or tens of thousands of images. Images of people comprise a significant portion of these images. Consumers capture images of the important people in their lives in a variety of social situations. People that are important to the photographer often appear many times throughout the personal collection. Many factors influence the position and pose of each person in the image. We propose that familial social relationships between people, such as "mother-child" or "siblings", are one of the strong factors. For example, Fig. 1 shows two images of a family at two different events. We observe that the relative position of each family member is the roughly the same. The position of a person relative to another is dependent on both the identity of the persons and the social relationship between them. To explore these ideas, we examine family image collections that have repeating occurrences of the same individuals and the social relationships that we consider are family relationships.

For family image collections, face recognition typically uses features based on facial appearance alone, sometimes including contextual features related to clothing [14,19,17]. In essence, that approach makes the implicit assumption that the identity of a face is independent of the position of a face relative to others

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part V, LNCS 6315, pp. 169–182, 2010.

Fig. 1. Social relationships often exhibit certain visual patterns. For the two people in a wife-husband relationship, the face that is higher in the image is more likely to be the husband. The family members are in roughly the same position in the two images, even though the images are of two different events on different days. The inclination of people to be in specific locations relative to others in a social relationship is exploited in this work for recognizing individuals and social relationships.

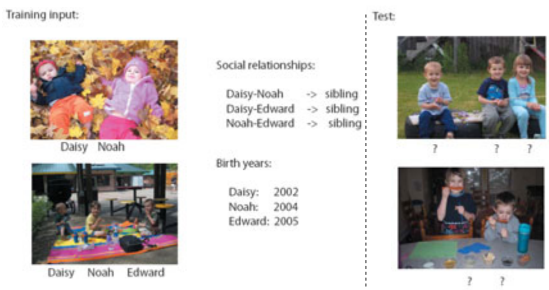


Fig. 2. In the training procedure, images are weakly labeled. Social relationships and birth years are annotated as input for learning social relationship models. In the recognition test procedure, the goal is to annotate faces present in images with names.

in the image. At its core, our work re-examines this assumption by showing that face recognition is improved by considering contextual features that describe one face relative to others in the image, and that these same features are also related to the familial social relationship.

Our contributions are the following: we develop a probabilistic model for representing the influence between pairwise social relationships, identity, appearance and social context. The experimental results show that adding social relationships results in better performance for face annotation. With the learned relationship models, we can in turn discover social relationships from new image collections where the social relationships are not manually annotated. To the best of our knowledge, this is the first work that shows that explicitly modeling social relationships improves person recognition. Further, this is the first work that demonstrates classification of social relationships from a single image. It is also important to note that our model is learned from an empirically attractive setting of weakly labeled data.

1.1 Related Work

Organizing consumer photo collections is a difficult problem. One effective solution is to annotate faces in photos and to search and browse images by people names [12]. Automatic face annotation in personal albums is a hot topic and attracts much attention [3,20]. There has been pioneering work on using social cues for face recognition [6,11,18]. [18] works with strongly labeled data, and only has one type of relationship: friend or not. In comparison, we deal with weakly labeled images, and explicitly model a number of social relationships. In [6], the authors uses the social attributes people display in pictures to better recognize genders, ages and identities. However, [6] does not explicitly model different social relationships between people or recognize specific individuals. In [11], recognizing individuals improves by inferring facial attributes. We extend these works by using social relationships as attributes for pairs of people in an image for recognizing people and social relationships.

Weak labeling is an area related to our work. In image annotation, ambiguous labels are related to generic object classes rather than names [1,8]. Berg et al. [2] is an example where face recognition has been combined with weak labels. In that work, face models are learned from news pictures and captions about celebrities, but ordinary people and the social relationships between them are not considered.

Certainly, the use of social relationships for recognition constitutes a type of context. The social context is related to the social interactions and environment in which an image is captured, and consequently it is not necessarily inferred directly from image data. Our contextual features for describing the relative positions between pairs of people in an image are similar to the contextual features shown to be effective in general object recognition [4,9,15]. In these works, pairwise features enforce priors that, for example, make it unlikely for cows to appear in the sky. We show that our similar features are in fact also useful for improving person recognition and for identifying social relationships. In our work, social relationships act as a high-level context leveraged from human knowledge or human behavior. In this sense, it is similar to the context of [5,16].

2 Approach

The common method for providing labeled samples to construct a model of facial appearance for a specific individual involves asking a user to label a set of training faces for each person that is to be recognized. Then, a face model can be learned in a fairly straightforward manner. However, annotating specific faces in a manual fashion is a time-consuming task. In practice, tools such as Flickr, or Adobe Album are used by many consumers, but they only provide weak labels that indicate the presence of a person but not that person's location in the image. Appearance models can still be learned in this scenario, but the label ambiguity increases the learning difficulty. In our work, we assume this realistic weak-labeling scenario, similar to that of [2], and our model is used to disambiguate the labels, learn appearance models, and find the identity of persons in images that were not in the original training subset. Note also that other frameworks exist for minimizing the effort of the user by using active learning to suggest samples to label [19,10], and our model could be inserted into one of these frameworks.

The procedure is illustrated in Fig. 2. For each image, we only know there are N names annotated, which are written as $\{p_i, i = 1, \cdots, N\}$, but do not know the positions or scales of the corresponding faces. Most of faces are automatically detected from images, and we manually add missed faces since we are not studying face detection in this work. Each face is represented by Fisher subspace features. Features of faces are written as $\{w_j, j = 1, \cdots, M\}$.

We train a face model for each individual. This requires establishing correspondences between names and faces in each training image. Social relationships are manually annotated by photo owners; the relationship between the i^{th} and j^{th} people is written as r_{ij}, a discrete variable over the nine pairwise social relationships that we consider. The labeling of this social relationship is reasonable and requires only a small amount of additional effort, because a given pairwise social relationship need be annotated only *once* for the entire personal collection. There are $N(N-1)/2$ possible pairwise relationships in one album with N people, but many pairs of people do not have direct relationships.

Table 1. The notation for our model

p_i: the i^{th} person name	P: all names
w_i: the feature representation of the i^{th} face	W: all face features
t_i: the age of the i^{th} person	T: all ages
r_{ij}: the social relationship between the i^{th} and the j^{th} person	R: all annotated relationships
f_{ij}: the social relationship features between the i^{th} and the j^{th} face	F: all social relationship features
A: the hidden variable which assigns names to faces	$A_i = j$: the i^{th} name is assigned to the j^{th} face
θ: model parameters	

A specific social relationship usually exhibits common visual patterns in images. For example, in a "husband-wife" relationship, the husband is usually taller than the wife due to physical factors (e.g., the average adult male is 176.8 cm while the average female is 163.3 cm [13]). Of course, it is easy to find exceptions, and this is why our model relies not on "rules" that define the behavior of an individual or a person in a family relationship, but rather on probabilistic distributions of features f for particular social relationships.

We extract features that reflect social relationships for each pair of faces i and j. The features describing the i^{th} and j^{th} face pair are written as f_{ij}. This feature vector represents the "social context" in our model. Note that even within a single social relationship, visual patterns are not time-invariant. For example, for "child-mother" relationship, when the child is an infant and the mother is in her 20s, the mother's face is physically larger than and generally positioned above the child's; but when the child grows, he or she may eventually have a larger face, be physically taller, and will no longer sit on the mother's lap. To accommodate the evolving roles within a social relationship, we allow the representation of social relationships for different age combinations. This requires that the collection owner provides approximate birth years for each person as illustrated in Fig. 2. In a training image, ages of people are written as $\{t_i, i = 1, \cdots, N\}$.

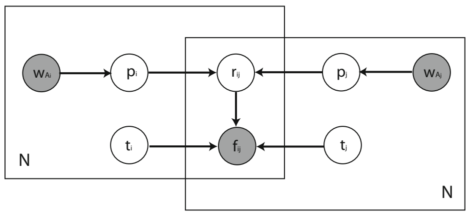


Fig. 3. The graphical model. The notation is explained in Table 1.

Given the above defined notations, we then aim to maximize the conditional probability of labels given image observations $p(P, R, T \mid W, F)$, which can be rewritten as:

$$\frac{p(P, R, T, W, F)}{p(W, F)} \sim \sum_A p(P, R, T, W, F \mid A)p(A) \tag{1}$$

A is a hidden variable that defines the correspondence between faces and names. $A_i = j$ denotes the i^{th} name is assigned to the j^{th} face. Given a specific A, the dependency between P, R, T, W and F is represented as shown in Fig. 3. We use a discriminative model to represent the appearance of each name (here we use a weighted KNN classifier due to its robustness, but note that a generative model such as a Gaussian mixture model is also applicable) and generative models for social relationships.

According to the graphical model, (1) can be written as:

$$\sum_A \prod_{i=1}^{N} p(p_i \mid w_{A_i}) \prod_{i=1,j=1}^{N} p(f_{A_i A_j} \mid r_{ij}, t_i, t_j)p(r_{ij} \mid p_i, p_j)p(A) \tag{2}$$

where w_{A_i} denotes the features of the face that is associated with the name p_i. r_{ij} is annotated for each pair of names p_i and p_j, so $p(r_{ij} \mid p_i, p_j)$ is 1 and neglected from now on. $p(p_i \mid w_{A_i})$ is calculated as:

$$p(p_i \mid w_{A_i}) = \frac{\sum_{l=1}^{L} p(p_i \mid w_l^{N_{A_i}})}{\sum_{i=1}^{N} \sum_{l=1}^{L} p(p_i \mid w_l^{N_{A_i}})} \tag{3}$$

Where $w_l^{N_{A_i}}$ denotes the l nearest neighbor faces found for w_{A_i} in all the training images. $p(p_i \mid w_l^{N_{A_i}}) = 0$ if the image containing $w_l^{N_{A_i}}$ does not have the person p_i present. $\sum_i p(p_i \mid w_j) = 1$ is enforced in the training procedure.

$f_{A_i A_j}$ denotes the social relationship features extracted from the pair of faces A_i and A_j. We extract five types of features to represent social relationships, which are introduced in Section 3. The space of each feature is quantized to several discrete bins, so we can model $p(f^k_{A_i A_j} \mid r_{ij}, t_i, t_j)$ as a multinomial distribution, where k denotes the k^{th} type of relationship features. For simplicity, these relationship features are assumed to be independent of each other, and $p(f_{A_i A_j} \mid r_{ij}, t_i, t_j)$ could simply be calculated as the product of the probability for each feature. However, we find that the features can be combined in smarter ways. By providing a learned exponent on each probability term, the relative importance of each feature can be adjusted. By learning the exponents with cross-validation on training examples, better performance is achieved.

There are many possible t_i and t_j pairwise age combinations, but we may only have a few training examples for each combination. However, visual features do not change much without a dramatic change of age. So we quantize each age t_i into 5 bins. The quantization partition points are $\left[0\ 2\ 17\ 35\ 60\ 100\right]$ years. Consequently, there are 25 possible pairwise age bin combinations. For each, we learn a multinomial distribution for each type of relationship feature. The multinomial distribution parameters are smoothed with a Dirichlet prior.

2.1 Learning the Model with EM

Learning is performed to find the parameters $\widehat{\theta}$:

$$\widehat{\theta} = \text{argmax}_\theta p(P, R, T \mid W, F; \theta) \tag{4}$$

θ contains the parameters to define $p(p \mid w)$ and $p(f \mid r, t)$. This can not be learned with maximum likelihood estimation because of the hidden variable. Instead, we use the EM algorithm, which iterates between the E step and the M step. Initialization is critical to the EM algorithm. In our implementation, we initialize $p(p_i \mid w_j)$ with the parameters produced by the baseline model that omits the social relationship variables. The multinomial distribution is initialized as a uniform distribution.

In the E step, we calculate the probability of the assignment variable A given the current parameters θ^{old}. For a particular A^*, we calculate it as:

$$p(A^* \mid P, R, T, W, F; \theta^{\text{old}}) = \frac{p(P, R, T, W, F \mid A^*; \theta^{\text{old}}) p(A^*; \theta^{\text{old}})}{\sum_A p(P, R, T, W, F \mid A; \theta^{\text{old}}) p(A; \theta^{\text{old}})} \tag{5}$$

$p(P, R, T, W, F, A^*; \theta^{\text{old}})$ can be calculated according to (2). The prior distribution of A is simply treated as a uniform distribution. This needs to be enumerated over all the possible assignments. When there are a large number of people in images, it becomes intractable. We only assign one p_i to a w_j when $p(p_i \mid w_j)$ is bigger than a threshold. In this way, we can significantly reduce the number of possible A.

In the M step, we update the parameters by maximizing the expected likelihood function, which can be obtained by combing (2) and (5). There are two types of parameters, one to characterize $p(p \mid w)$ and the other one to characterize $p(f \mid r, t)$. In the M step, when updating one type of parameters using maximum likelihood estimation, the derivative doesn't contain the other type of parameters. Therefore, the updates of parameters for $p(p \mid w)$ and $p(f \mid r, t)$ are separate. When running the EM algorithm, the likelihood values do not change significantly after 5 to 10 iterations.

2.2 Inference

In the inference stage, we are given a test image containing a set of people (without any name label information), we extract their face appearance features W and relationship features F, then predict the names P. We use the relationship models to constrain the labeling procedure, so the classification of faces is not done based on facial appearance alone. This problem is equivalent to finding a one-to-one constraint $A^\star$ in the following way:

$$A^\star = \text{argmax}_A p(A \mid P, R, W, F, T) \tag{6}$$

Here, P denotes all the names in the dataset. There would be too many possible A to evaluate and compare. We adopt a simple heuristic by only considering As which assign a name p to a face w when $p(p \mid w)$ is bigger than a threshold. This heuristic works well in our implementation.

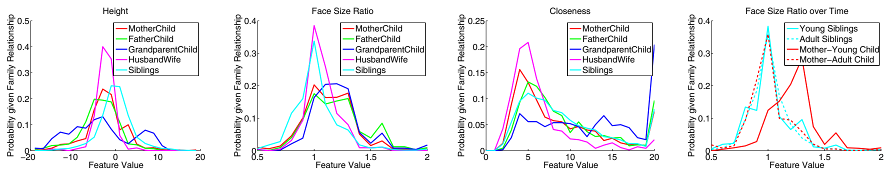
3 Implementation Details

In this section, we describe important implementation details. The appearance of each face is represented by projecting the original pixel values into a Fisher subspace learned from a held-out collection (containing no images in common with either the training set or the test set). Each face is represented as a Fisher discriminant space feature.

In our model, the social relationship variable r_{ij} is discrete over the space of pairwise social relationships. We represent the following nine familial social relationships between a pair of people:

mother-child	father-child	grandparent-child	husband-wife	siblings
child-mother	child-father	child-grandparent	wife-husband	

We consider relationships to be asymmetric (e.g., “mother-child” is different from “child-mother”) because our objective is to identify the role of each person in the relationship. We use the following five types of observed appearance features to represent social relationships.



(a) Height (b) Face Size Ratio (c) Closeness (d) Size over time

Fig. 4. Pairwise facial features are dependent on social relationships. From these plots, we see that parents' faces are usually above childrens' faces (a), that spouses' faces are usually about the same size, but are larger than children's (b), and spouses tend to be close together in an image(c). Note that we also model the changing nature of family relationships over time: a mother's face is larger than the child's when the child is young, but they are generally the same size when the child is an adult (d).

Height: the height difference is used as a feature. Very simply, we use the ratio of the difference y-coordinates of the two people's faces to the average face size of the faces in the image. The ratio is quantized to six bins.

Face size ratio: this feature is the ratio of the face sizes. We quantize the ratio to six bins.

Closeness: the distance of two people in an image can reveal something about their social relationship. We calculate the Euclidean distance between pair of people, normalized by the average face size. We quantize the distance to five bins.

We train gender and age classifiers based on standard methods, following the examples of [7,11]. Two linear projections (one for age and one for gender) are learned and nearest neighbors (using Euclidean distance) to the query are found in the projection space.

Age difference: we use our age predictor to estimate the ages of people. This age difference, estimated purely from appearance, tells us some information about the social relationship. We quantize age into five ranges, so the age difference between two people has nine possibilities. The age difference relationship is modeled as a multinomial distribution over these nine bins.

Gender distribution: the appearance-based gender classifier helps to indicate the role of a person in a social relationship. For example, gender estimates are useful for distinguishing between a wife and husband (or more broadly a heterosexual couple). For each pair of people, there are four possible joint combinations of the genders.

Fig. 4 demonstrates evidence of the dependence between social relationships and our features by showing the distribution of feature values given the social relationships, as learned from our training collections.

4 Experiments

In this section, we show experiments that support our assertion that modeling social relationships provides improvements for recognizing people, and allows for the recognition of pairwise social relationships in new images.

In Section 4.1, we examine the task of identifying people through experiments on three personal image collections, each of which has more than 1,000 images and more than 30 distinct people. We show that significant improvement is made by modeling social relationships for face annotation on both datasets. We also investigate how different social relationships features help to boost the performance.

Furthermore, in Section 4.2, we show that learned social relationships models can be transferred across different datasets. Social relationships are learned on a personal image collection, and then social relationships are effectively classified in single images from unrelated separate image collections.

4.1 Recognizing People with Social Relationships

In the first experiment, a subset of images from a personal image collection is randomly selected as training examples, and weak name labels are provided for the identities of the people in the images. The remaining images comprise a test set for assessing the accuracy of recognizing individuals. Testing proceeds as follows: First, the correspondence between the names and the faces of the training images are found using the EM procedure from Section 2.1. Next, inference is performed (Section 2.2) to determine the most likely names assignment for each set of faces in each test image. The percentage of correctly annotated faces is used as the measure of performance. This measure is used to evaluate the recognition accuracy in the test set as well as in the training set.

The first collection has 1,125 images and contains 47 distinct people. These people have 2,769 face instances. The second collection contains 1,123 images, with 34 distinct people and 2,935 faces. The third collection has 1,117 images of 152 individuals and 3,282 faces. For each collection, we randomly select 600 images as training examples and the others as test examples. Each image contains at least two people. In total, these images contain 6,533 instances of 276 pairwise social relationships.

Improvement made by modeling social relationships: For comparison to our model that includes social relationships, we first perform experiments without modeling social relationships. In the training procedure, we maximize: $p(P \mid W) \sim \sum_A \prod_{i=1}^N p(p_i \mid w_{A_i})p(A)$. Likewise, the EM algorithm is employed to learn model parameters.

Fig. 5 shows that all datasets show improved recognition accuracy in both training and testing when social relationships are modeled. By modeling social relationships, better correspondence (i.e. disambiguation of the weak label names) in the training set is established. In collection 1, training set accuracy improves by 5.0% by modeling social relationships, and test set identification improves by 8.6% due to the improved face models as well as the social relationship models. Significant improvement is also observed in collection 2 in both the training (improves by 3.3%) and test (improves by 5.8%) sets. Collection 3 also shows improvement (by 9.5% in training and by 1.8% in testing) although the overall accuracy is lower, mainly because this collection contains many more unique people (152 people versus 47 and 34 in collections 1 and 2).

Fig. 6 illustrates the improvement that modeling social relationships provides for specific test image examples. The faces in green squares are instances that are not correctly classified when the model ignores social relationships, but are corrected by modeling social relationships. We can see that these faces are surrounded by other people who have strong social relationships with, and the visual patterns between people are what is typically expected given their roles in the relationships. The faces in red squares are instances that are correctly classified when appearance alone is considered, but get confused by incorporating social relationships. This is because visual relationship patterns in these pictures are atypical of what is observed in most of other pictures. mother, so she is misclassified as her father, despite her childlike facial appearance.

Table 2. Person recognition accuracy in the test set improves for both collections by modeling social relationships using more features. For example, "+height" means that only relative height feature is used, and the other features are omitted.

	without relationships	+height	+closeness	+size	+age	+gender	+all
Collection 1	0.560	0.621	0.628	0.637	0.635	0.630	**0.646**
Collection 2	0.537	0.563	0.560	0.583	0.573	0.584	**0.595**
Collection 3	0.343	0.361	0.359	**0.362**	0.362	0.362	0.361
Overall Mean	0.480	0.515	0.516	0.527	0.523	0.525	**0.534**

Effect of each social relationship feature: As described in Section 3, we use five features to encapsulate social relationships. We show how each type of relationship feature helps by in turn omitting all features except that one. The results are shown in Table 2. We observe that relative face size is the most helpful single feature, followed by age and gender. In general, including all features provides significant improvement over using any single feature and adding any single feature is better than using none at all. It is interesting to note that while our results concur with [11] in that we achieve improved face recognition by estimating age and gender.

4.2 Recognizing Social Relationships in Novel Image Collections

Our model explicitly reasons about the social relationships between pairs of people in images. As a result, the model has applications for image retrieval based on social relationships.

Social relationships are modeled with visual features such as relative face sizes and age difference, which are not dependent on the identities of people. This

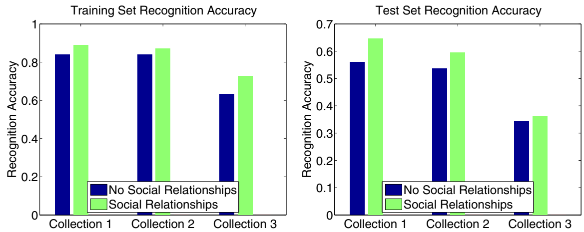
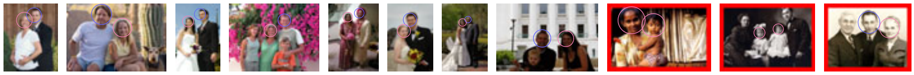


Fig. 5. Modeling social relationships improves recognition accuracy. The plots show the improvement in recognition accuracy for both the training set (left) and the test set (right) for two different image collections.

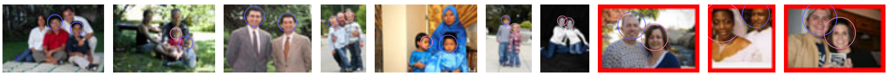
Fig. 6. The faces in green squares are instances that are not correctly recognized without modelling social relationships, but are corrected by modeling social relationships. The faces in red squares are correctly recognized at first, but are misrecognized when social relationships are considered. The mistakes are sometimes due to an improbable arrangement of the people in the scene (e.g. the son on the father's shoulders in the lower right) that is not often observed in the training set. As another example, in the middle image of the second row, the daughter (closer to the camera) appears taller and has a bigger face size than her mother, so she is misclassified as her father, despite her childlike facial appearance.

means social relationship models can be transferred to other image collections with different people. Consequently, the models learned from one image collection can be used to discover social relationships in a separate unrelated image collection with no labeled information at all. We perform two experiments to verify that we learn useful and general models for representing social relationships in images.

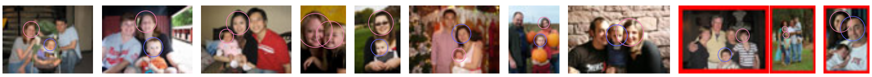
In the first experiment, we learn social relationship models from the training examples of collection 1, and classify relationships in collection 2. Because collection 2 contains no "grandparent-child" relationships, we limit the classified r_{ij} values to the other seven social relationships. The confusion matrix is shown in Fig. 8. Each row of this confusion matrix shows an actual class.



(a) social relationships classified as wife-husband



(b) social relationships classified as siblings



(c) social relationships classified as mother-child

Fig. 7. Social relationship classification is accomplished from single images with our model, trained only with weak labels on a single, unrelated personal collection. Here, the task is to distinguish between the "wife-husband", "siblings", and "mother-child" relationships for each pair of circled faces. Incorrect classifications are outlined in red.

	child–mother	mother–child	child–father	father–child	wife–husband	husband–wife	sibling
child–mother	.71	.01	.24			.01	.03
mother–child	.01	.71	.01	.24			.03
child–father	.39	.01	.32		.01	.23	.04
father–child	.01	.38		.33	.24	.01	.04
wife–husband	.07		.04	.01	.35	.05	.48
husband–wife		.07	.01	.03	.04	.40	.44
sibling	.07	.07	.02	.02	.05	.05	.73

	child–mother	mother–child	wife–husband	husband–wife	sibling
child–mother	.52		.18	.09	.21
mother–child	.01	.65	.04	.25	.05
wife–husband	.15	.08	.52	.17	.08
husband–wife	.08	.16	.04	.53	.19
sibling	.06	.12	.15	.24	.42

Fig. 8. The confusion matrix of social relationships classification. **Left:** We learn social relationship models from collection 1 and test on the images of collection 2. **Right:** We apply the learned social relationship models to a set of images from Flickr, and labeled as one of five social relationships. Both experiments show that social relationship models learned from one collection and transferable and useful for classifying social relationships in images containing strangers.

The averaged value of diagonals is 50.8%, far better than random performance (14.3%). We can see that the mistakes are reasonable. For example, "child-mother" is usually misclassified as "child-father" because the primary visual difference between "mother" and "father" is the gender, which may not be reliably detected from consumer images.

In a second experiment, we perform social relationship recognition experiments on the publicly released group image dataset [6]. First, we manually labeled relationships between pairs of people. A total of 708 social relationships were labeled, at most one relationship per image, and each of the three social relationships has over 200 samples. This dataset is used solely as a test set. The social relationship models are learned from collection 1 in the same weakly supervised learning fashion as before. The confusion matrix is shown in Fig. 8. The overall social relationship classification accuracy in this experiment is 52.7%, again exceeding random classification 20.0%. This performance is significant in that the entire model is trained on a single personal image collection with weak labels. Images classification results from the model are shown for three social relationships in Fig. 7.

5 Conclusions

We introduce a model that incorporates pairwise social relationships such as husband-wife or mother-child for representing the relationship between people in a personal image collection. This model is motivated by the observation that the joint appearance between people in an image is associated with both their identities and the social relationship between the pair. We show experimentally several advantages of this representation. First, the model allows for establishing the correspondence between faces and names in weakly labeled images. Second, the identification of unknown faces in test images is significantly improved when social relationship inference is included. Third, social relationships models learned from the weakly labeled data are used to recognize social relationships in single previously unseen images. This work is believed to represent the first attempt at explicitly modeling the pairwise social relationships between people in single consumer images.

Acknowledgement. This work was supported in part by the National Science Foundation under IIS -0803603 and in part by the Office of Naval Research under N00014-01-1-0890 as part of the MURI program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation or the Office of Naval Research.

References

1. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D.M., Jordan, M.I.: Matching words and pictures. JMLR 3, 1107–1135 (2003)
2. Berg, T., Berg, A., Edwards, J., Maire, M., White, R., Teh, Y.W., Learned-Miller, E., Forsyth, D.: Names and faces in the news. In: Proc. CVPR (2004)
3. Chen, L., Hu, B., Zhang, L., Li, M., Zhang, H.J.: Face annotation for family photo album management. IJIG 3(1), 81–94 (2003)
4. Divvala, S.K., Hoiem, D., Hays, J., Efros, A., Hebert, M.: An empirical study of context in object detection. In: Proc. CVPR (2009)

5. Gallagher, A., Chen, T.: Estimating age, gender, and identity using first name priors. In: Proc. CVPR (2008)
6. Gallagher, A., Chen, T.: Understanding Images of Groups of People. In: Proc. CVPR (2009)
7. Guo, G., Fu, Y., Dyer, C., Huang, T.: Image-based human age estimation by manifold learning and locally adjusted robust regression. In: IEEE Trans. on Image Proc. (2008)
8. Gupta, A., Davis, L.S.: Beyond nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 16–29. Springer, Heidelberg (2008)
9. Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. IJCV 80(1), 3–15 (2008)
10. Kapoor, A., Hua, G., Akbarzadeh, A., Baker, S.: Which Faces to Tag: AddingPrior Constraints into Active Learning. In: Proc. ICCV (2009)
11. Kumar, N., Berg, A., Belhumeur, P., Nayar, S.: Attribute and simile classifiers for face verification. In: Proc. ICCV (2009)
12. Naaman, M., Yeh, R., Garcia-Molina, H., Paepcke, A.: Leveraging context to resolve identity in photo albums. In: Proc. JCDL (2005)
13. National Center for Health Statistics. CDC growth charts, United States (2007), `http://www.cdc.gov/nchs/data/nhanes/growthcharts/zscore/statage.xls`
14. O'Hare, N., Smeaton, A.: Context-aware person identification in personal photo collections. IEEE Trans. MM (2009)
15. Parikh, D., Zitnick, L., Chen, T.: From appearance to context-based recognition: Dense labeling in small images. In: Proc. CVPR (2008)
16. Pellegrini, S., Ess, A., Schindler, K., van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: Proc. ICCV (2009)
17. Song, Y., Leung, T.: Context-aided human recognition- clustering. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 382–395. Springer, Heidelberg (2006)
18. Stone, Z., Zickler, T., Darrell, T.: Autotagging Facebook: Social network context improves photo annotation. In: Proc. CVPR Internet Vision Workshop (2008)
19. Tian, Y., Liu, W., Xian, R., Wen, F., Tang, X.: A face annotation framework with partial clustering and interactive labeling. In: Proc. CVPR (2007)
20. Zhao, M., Teo, Y.W., Liu, S., Chua, T., Jain, R.: Automatic person annotation of family photo album. In: Sundaram, H., Naphade, M., Smith, J.R., Rui, Y. (eds.) CIVR 2006. LNCS, vol. 4071, pp. 163–172. Springer, Heidelberg (2006)