KINSHIP CLASSIFICATION BY MODELING FACIAL FEATURE HEREDITY

Ruogu Fang[1], *Andrew C. Gallagher*[1], *Tsuhan Chen*[1], *Alexander Loui*[2]

[1]Dept. of Elec. and Computer Eng., Cornell University [2]Eastman Kodak Company

ABSTRACT

We propose a new, challenging, problem in kinship classification: recognizing the family that a query person belongs to from a set of families. We propose a novel framework for recognizing kinship by modeling this problem as that of reconstructing the query face from a mixture of parts from a set of families. To accomplish this, we reconstruct the query face from a sparse set of samples among the candidate families. Our sparse group reconstruction roughly models the biological process of inheritance: a child inherits genetic material from two parents, and therefore may not appear completely similar to either parent, but is instead a composite of the parents. The family classification is determined based on the reconstruction error for each family. On our newly collected "Family101" dataset, we discover links between familial traits among family members and achieve state-of-the-art family classification performance.

Index Terms— kinship classification, facial inheritance, sparsity, sparse group lasso, family

1. INTRODUCTION

Kin recognition is a new research arena in computer vision that has attracted increasing attention in recent years [1, 2, 3]. In computer vision, kin recognition is the task of training the machine to recognize the genetic kin and non-kin based on features extracted from digital images. In this paper, we are interested in the problem of kinship family classification: Given a set of families, each with a set of images of family members, the goal is to determine which family that a person in a query image belongs to. We present the idea that kinship classification from facial images is improved when we use sparsity to model the genetic process of heredity of visible traits.

The computer vision facial representations do not answer questions about why people look the way that they do, or how a woman's face's appearance is affected by her ancestors. To answer questions along this vein, we must appeal to the field of genetics. Based on work from the 1860's, Gregor Mendel is credited with discovering the laws of the inheritance of visible traits between generations of offspring (specifically pea plants) in sexual reproduction. Mendel produced offspring from various pea plants with different visible traits (e.g. white or purple flowers), and observed the proportion of offspring with each traits. From this evidence, he proposed the idea of factors (i.e., alleles of a gene) that code for either dominant or recessive traits.

Our main contributions are the following: First, we propose a sparse group lasso that models the inheritance of facial features (Fig. 1). We show that our model performs well at a new kinship classification task: determining which family from a set of families that a query person belongs to. Finally, we have created and will share a large database of families, named "Family101", which contains 101 public families with cross-generation family structure, with 607 individuals and 14,816 facial images.

Fig. 1. We use facial parts of family members to reconstruct a query person. Given a large number of facial parts from various families as dictionary, as shown above, a query person's image can be reconstructed from a sparse group of facial part bases from the dictionary. The height of a bar reflects the importance of this base in reconstructing the query person. Interestingly, the non-zero weights represent the images of people from Obama's family, while people from other families get zero weights.

2. RELATED WORK

We review related work on sparsity, face recognition, and kinship recognition.

Sparse group lasso: Group sparsity was introduced in [4, 5]. For a given a test image, the non-zero linear reconstruction coefficients should ideally only appear in one or a few classes. It aims to select only a few classes to reconstruct the test data. To emphasize the sparse characteristic on both the instance level and the group level, sparse group lasso is a combination of group sparse coding and sparse coding. Group sparsity has been applied to image annotation [6] and region labeling [7]. Our work learns the sparse coefficients with respect to the facial parts from the family dictionary, and the resulting reconstruction residues from all parts are used to identify the family and also relatives who have similar facial traits.

Face recognition: Numerous methods have been proposed to tackle this long-standing challenge, including Eigenfaces [8], Fisherfaces [9] and SVM [10]. Recently, Wright et al. [11] applied a sparse representation-based classification (SRC) approach to frontal view face recognition with varying expression and illumination, as well as occlusions and disguise, and achieved impressive performance, which was followed by several extensions [12]. All these methods are restricted to face recognition problem, not kinship recognition, and enforce sparsity on the entire dictionary, instead of group sparsity which enforces additional sparsity on the group level.

Work on face verification addresses the enormous variability in which the same face presents itself to a camera. Kumar et al. [13] explore face verification using attribute and simile classifiers. Yin et

Fig. 2. An example of kinship classification task between two possible families with both candidate parents displayed. A direct one-to-one appearance-based approach might classify the query to family (a) due to the similarity between the query person and the mother. However, the actual related family is (b). The query person is Shannon Lee, daughter of Bruce Lee and Linda Lee Cadwell.

al. [14] also achieves good verification accuracy using an associate-predict model. These works, although related, are tackling the problem of verifying the same face with variability caused by pose, expression, hairstyle, etc. In contrast, our work involves different faces from the same family, and therefore requires a new framework by reconstructing the query face from a mixture of parts from a set of family members

Kinship recognition: Human perception of kinship is an active research area in psychology [15, 16, 17, 18, 19, 20]. The findings from the psychological literature include that (1) humans are able to recognize kinship even on unfamiliar faces; and (2) the mechanism of kinship perception is probably different from identity recognition. Motivated by the psychological findings, several attempts have been made to develop computational approaches to kinship recognition. Fang et al. [1] tackles the kinship verification task by extracting discriminative features from facial images and using the absolute difference between pairs of people to determine the existence of kinship. Xia et al. [2] utilizes young parents as an intermediate distribution to bridge the children and the old parents and reduce the divergence caused by age via transfer learning. Unlike our approach, these methods tackle the one-to-one verification problem without considering feature inheritance from multiple relatives, and do not model facial parts specifically. In contrast, our kinship classification model is one-to-multiple that better resembles the biological process of inheritance, as illustrated in Fig. 2. Our method models the feature inheritance across multiple individuals, and it is also able to estimates which relatives have similar facial features to the query person.

3. APPROACH

We segment the query face into parts (eyes, nose, etc.) and reconstruct each part as a linear combination of a set of database parts. We assume that we have a training set of images of different families, and within each family there are several images of the same person.

3.1. Notation

The traditional Lasso problem is formulated as:

$$\min_{\alpha} \|y - D\alpha\|_2^2 \quad s.t. \|\alpha\|_1 \leq \sigma \tag{1}$$

where $\sigma > 0$ is a constant, $y \in \mathbf{R}^n$ is the signal to be coded, $D = [d_1, \ldots, d_k] \in \mathbf{R}^{n \times k}$ is the dictionary with column vector $d_j \in \mathbf{R}^n$ as the j^{th} atom, and $\alpha \in \mathbf{R}^k$ is the coding coefficient vector.

In our problem of kinship classification, y is the feature vector computed from the query person's image, D is the dictionary composed of the training images from various families, and α is a sparse coefficient vector to represent y with a combination of vectors from D. The query person does not appear in any base of the dictionary D, while the family relatives of the query person are in D.

Fig. 3. Twelve facial parts are selected to build the part-based dictionaries.

3.2. Part-based Dictionary

We use parts of the face instead of the entire face to address the kinship classification problem and model the process of facial feature heredity. We segment the facial image into 12 parts, following the approach in [13], as shown in Fig. 3. The location and the size of the 12 facial parts are determined from the annotation of one researcher on a template face and applied to all the other facial images which have been aligned to the same template. For each part p, we build a dictionary $D^{(p)}$ from the training samples. The atom $d_j^{(p)}$ is the feature extracted from the part p of the training face sample j in the dictionary $D^{(p)}$. This enables us to find the features in the dictionary that best reconstruct each part in the query face, and each part is the dynamic combination of the features from the part-based dictionary.

3.3. Kinship Classification with Sparse Group Lasso

We formulate the problem as a sparse group lasso, where we enforce sparsity on both the group (family) and individual feature levels. This is based on the observation that the query person inherits features from related family members, thus the non-zero coefficients in α tend to be in groups and indicate kinship. We only need a few images of the related family members with similar expression, illumination and pose to represent the features in the query image. By taking advantage of the coefficients' group structure, the performance of kinship classification model can be improved. The problem is formulated as:

$$\arg\min_{\alpha} \|y - D\alpha\|_2^2 + \lambda_1 \sum_{j=1}^{m} \|\alpha_{G_j}\|_2 + \lambda_2 \|\alpha\|_1 \tag{2}$$

where the training images are partitioned into m disjoint groups $G_1, G_2, \ldots, G_m$. $\alpha = [\alpha_{G_1}, \alpha_{G_2}, \ldots, \alpha_{G_m}]$ and α_{G_j} denotes the group of weights corresponding to group G_j. In the kinship classification problem, the groups are naturally defined as the different families linked by blood relation. $\lambda_1 \sum_{j=1}^{m} \|\alpha_{G_j}\|_2$ is the combination of L_1 and L_2 norms. The L_2 norm is used for the weights inside the same group, and the L_1 norm is used to sum the results between groups. Using the group structure ensures more robust and accurate weights and still benefits from the sparsity. $\lambda_2 \|\alpha\|_1$ is the L_1 norm to enforce the sparsity on the coefficients in α, which benefits from the sparsity by selecting only a few images with similar illumination, pose and expression.

In Eq. 2, the sparse factorization coefficient α indicates the distribution of y on all the bases in D, each of which encodes a specific feature vector of the candidate family member. A child inherits genetic material from two parents, and therefore may not appear completely similar to either parent. Ideally, the person that corresponds

to the same family should have high coefficients from related family members and low or zero coefficients from unrelated persons. To classify the family that a query person belongs to, we reconstruct the facial part p of the query face y, $y^{(p)}$, with the bases for facial part p from the family j, $D_j^{(p)}$, and the weights in α that correspond to family j, $\alpha_{G_j}^{(p)}$. The reconstruction error of facial part p for family j is $R_j^{(p)} = \|y^{(p)} - D_{G_j}^{(p)} \alpha_j^{(p)}\|_2^2$. R is then normalized by the size of the facial part.

From the genetic perspective, not all traits inherited from the parents to the child are dominant. The recessive trait may be expressed as a phenotype if the child inherits two copies of the gene from the parents. In this case, the child may have some facial traits distinct from both parents. So the reconstruction error of these parts will be large even for the related family. On the other hand, some dominant traits are representative for the family. A few characteristic traits passed from the parents to the child could help people to identify the kinship between the people easily. Thus, instead of using all the facial parts to classify kinship, we choose the three facial parts of the query person with the smallest possible residues among all $R_j^{(p)}, j = 1, 2, \ldots, k$. We rank the normalized reconstruction error for all candidate families for these three facial parts, and sum the ranks for each family. The kin related family is decided as the one with the smallest rank sum.

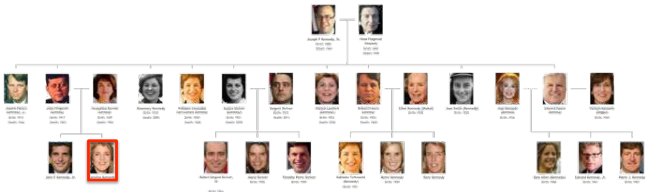
3.4. Familial Traits Tracing

By learning the sparse coefficient $\alpha^{(p)}$ from the part-based dictionary $D^{(p)}$ for each query image, we determine which family members have similar facial parts. We first find the family that the query person most likely belongs to by computing the rank-sum of the three facial parts with least reconstruction error as described in Section 3.3. Then we pick up the top three images in the part-based dictionary $D^{(p)}$ with the largest coefficients. These images are regarded as the images of family relatives who have similar facial traits.

4. DATASET

We assemble a new dataset called "Family 101" containing 101 different family trees, including 206 nuclear families, 607 individuals, with 14,816 images. The dataset includes renowned public families. The structure of the Family 101 dataset is illustrated in Fig. 4. To the best of our knowledge, it is the largest dataset with structured familial relationship. We will publish our dataset upon the publication of the paper.

We used Amazon Mechanical Turk to assemble the dataset by asking workers to upload images of family members that we specify. The identities of the individuals are then verified. Each family contains 1 to 7 nuclear families. In total there are 206 nuclear families (both parents and their children), each with 3 to 9 family members. The final dataset includes around 72% Caucasians, 23% Asians, and 5% African Americans to guarantee a widespread distribution of facial characteristics that depend on race, gender, age. We attempted to exclude non-biologically related parents-children by checking the familial relationships using public information available online. For pair-wise relationships, there are 213 father-son relations, 147 father-daughter relations, 184 mother-son relations, and 148 mother-daughter relations. For some of the family members in Family 101 dataset the Internet search returns relatively few unique photographs. For all experiments in this paper we use only 546 family members for which there are at least 5 unique photographs. The comparison between our dataset and the existing family image datasets are summarized in Table 1. Next, a commercial face detection package was used find faces six fiducial points per face: the corners of both eyes and the corners of the mouth, used for facial alignment.



(a) Kennedy family tree from FamNet dataset

(b) All 48 images of Caroline Kennedy

Fig. 4. Our Family 101 database contains 14,816 images of 607 public figures from 206 nuclear families from 101 renowned root families. Here, we show one of the 101 families: the Kennedy family tree with three generations. The variability in appearance captured by Family 101 can be seen in (b), which shows all 48 images of Caroline Kennedy.

Table 1. Comparison of our Family 101 dataset and other existing family image datasets of human faces. "Age" variation refers to the variation of age of the same individual, e.g. people when they are young and old. "Family structure" refers to the existence of cross-generation and family tree relationship in the dataset. Bold font indicates relatively larger scale datasets or more information.

Dataset	No. Family	No. People	No. Images	Age varies?	Family structure?
CornellKin[1]	150	300	300	No	No
UB KinFace[2]	90	180	270	Yes	No
Family 101	206	607	14,816	Yes	Yes

5. EXPERIMENTS

5.1. Experiment Setup

We evaluate our approach on the "Family101" dataset. We use a dense SIFT descriptor on rescaled facial images of size 61×49 pixels as our image feature. The SIFT descriptors extracted from 16×16 pixel patches are densely sampled from each image on a grid with step size 6 pixels. The images were all preprocessed to gray scale. The parameters controlling the sparsity of the coefficients λ_1 and λ_2 are determined on a separate validation dataset of 20 families which do not overlap with the testing data. The parameters are set as $\lambda_1 = 0.1, \lambda_2 = 0.1$.

We compare our part-based reconstruction to three baselines:

Baseline: Our first two baselines are the widely used discriminative classifiers, K-nearest-neighbors and support vector machine. For these two baselines, we tune the parameters to reach the best possible performance for kinship classification. Our third baseline sparse representation based recognition (SRC) proposed by Ma et

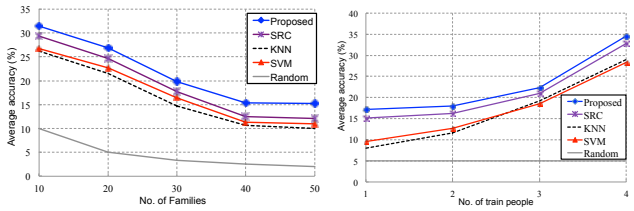


Fig. 5. Left: Our proposed sparse reconstruction method is the most robust at classifying the family that a query face belongs to. When there are more families to choose from, overall performance decreases, but the improvement margin of our proposed method increases. Right: Performance improves as the number of training exemplars per family increases.

al. [11] for robust face recognition, which uses L_1 regularization on the whole face. All four classification algorithms (3 baselines and our proposed method) use the same features, i.e. the dense SIFT features described above, for fair comparison. SRC in our baseline also uses the dense SIFT feature instead of the pixel intensity, but the classification framework is same as [11].

Setup: We compare all methods in several different scenarios. *Unless specified in each scenario*, for each family, we use 3 family members for training (i.e., the training individuals comprise the sparse dictionary) and 2 for testing. 20 families are randomly selected from the Family101 dataset for evaluation in every scenario. Families with fewer than 5 family members are not used. We use 30 images of each individual for both training and testing, and report mean per-family accuracy over 10 random family selected for training and the remaining 10 for testing.

5.2. Experimental Results

We validate our propose sparse group lasso method for kinship classification in three different scenarios, by tuning the number of families for evaluation, number of family members for training and tracing the facial traits.

Number of families: We first study kinship classification as the number of families (or categories) increases. Fig. 5 (left) shows the results where our proposed method surpasses the baselines for the problem of kinship classification.

First, we see that all of the four methods perform much better than chance, while our proposed method consistently performs better than other three algorithms. This validates the power of part-based reconstruction. In addition, SRC and our proposed methods gains over KNN and SVM with larger number of families demonstrate the benefit of sparse reconstruction modeling.

As one would expect, accuracy for all four approaches decreases with more families in the gallery. Our method surpasses the baselines for the whole range, and the gain of our method compared to the baselines grows as the number of families increase.

Number of family members for training: We next study the impact of varying number of family members for training. How many members from one family does the computer need to see before it can classify an unknown person to one of the families with reasonable accuracy? We keep the total number of family members same for each family, while varying the number of family members who are to be classified. Fig. 5 (right) shows the results as we increase the number of family members for training, where for each family, we randomly select n training individuals from available family member, and then use $5 - n$ for testing, to keep the total number of family members for both training and testing the same.

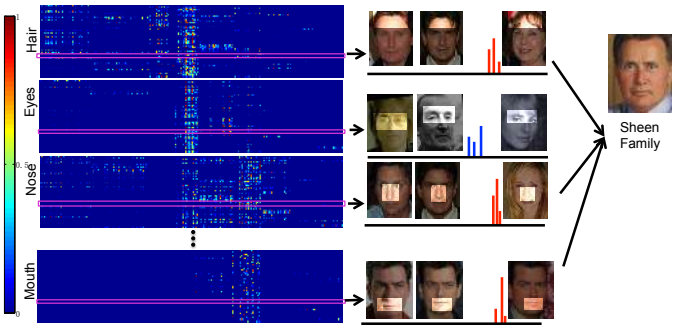


Fig. 6. Visualization of the sparse representation from a family facial parts. Each row in the left-most matrices corresponds to one reconstruction coefficient for a test image with respect to 50 people from 10 families. Red indicates large magnitude in the coefficient while blue indicates low magnitude. We observe that the bases are indeed very sparse. For a query image of Martin Sheen, the three parts with the least reconstruction errors are hair, eyes and mouth. In the middle column, red bars indicate reconstruction from related family members, while the blue bars indicate reconstruction from unrelated people. We show the actual images that are used to reconstruct the parts above the axis. Typically, reconstruction error is higher when a part is reconstructed from unrelated people (versus family members).

For all approaches, performance improves as the number of training people increases. Our proposed method performs better than the baselines over most of the spectrum.

When using more than two individuals for training, our method's advantage of utilizing the combinatorial information from multiple family members is shown by the faster increase of accuracy, which makes sense because the children inherit features not only from one parent or a single relative family member, and share features in common with other relatives (besides parents).

Facial traits tracing: A visualization of the sparse coefficients of query images with respect to a set of families is shown in Fig. 6. We observe that non-zero coefficients tend to cluster in groups, and typically come from individual family members who have similar facial traits with the query person. However, due to the noise caused by various factors and potential similarity between people who are not related by family linkage, e.g., Martin Sheen may has similar eyes as Cynthia Lennon in this photo, but most of non-zero coefficients (e.g., for hair, nose, mouth) are from the Sheen family.

6. CONCLUSION

We propose a novel framework for recognizing kinship by modeling this problem as that of reconstructing the query face from a mixture of parts from a set of families, which allows for a richer utilization of family members' information and family structure than the commonly one-to-one kinship verification. Our model is motivated by the biological process of inheritance. Various experiments on our newly collected dataset "Family101" demonstrate the advantages of our idea.

7. REFERENCES

[1] Ruogu Fang, Kevin D Tang, Noah Snavely, and Tsuhan Chen, "Towards computational models of kinship verification," in

Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE, 2010, pp. 1577–1580.

[2] Siyu Xia, Ming Shao, and Yun Fu, "Kinship verification through transfer learning," in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three*. AAAI Press, 2011, pp. 2539–2544.

[3] Siyu Xia, Ming Shao, Jiebo Luo, and Yun Fu, "Understanding kin relationships in a photo," *Multimedia, IEEE Transactions on*, vol. 14, no. 4, pp. 1046–1056, 2012.

[4] Samy Bengio, Fernando Pereira, Yoram Singer, and Dennis Strelow, "Group sparse coding," *Advances in Neural Information Processing Systems*, vol. 22, pp. 82–89, 2009.

[5] Xiao-Tong Yuan and Shuicheng Yan, "Visual classification with multi-task joint sparse representation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3493–3500.

[6] Shaoting Zhang, Junzhou Huang, Yuchi Huang, Yang Yu, Hongsheng Li, and Dimitris N Metaxas, "Automatic image annotation using group sparsity," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3312–3319.

[7] Xiaobai Liu, Bin Cheng, Shuicheng Yan, Jinhui Tang, Tat Seng Chua, and Hai Jin, "Label to region by bi-layer sparsity priors," in *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 2009, pp. 115–124.

[8] Matthew Turk and Alex Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[9] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[10] Bernd Heisele, Purdy Ho, and Tomaso Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 688–694.

[11] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[12] Yanan Liu, Fei Wu, Zhihua Zhang, Yueting Zhuang, and Shuicheng Yan, "Sparse representation using nonnegative curds and whey," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3578–3585.

[13] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar, "Attribute and simile classifiers for face verification," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 365–372.

[14] Qi Yin, Xiaoou Tang, and Jian Sun, "An associate-predict model for face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 497–504.

[15] Maria F Dal Martello and Laurence T Maloney, "Where are kin recognition signals in the human face?," *Journal of Vision*, vol. 6, no. 12, 2006.

[16] Maria F Dal Martello and Laurence T Maloney, "Lateralization of kin recognition signals in the human face," *Journal of vision*, vol. 10, no. 8, 2010.

[17] Alexandra Alvergne, Charlotte Faurie, and Michel Raymond, "Father–offspring resemblance predicts paternal investment in humans," *Animal Behaviour*, vol. 78, no. 1, pp. 61–69, 2009.

[18] Lisa M DeBruine, Finlay G Smith, Benedict C Jones, S Craig Roberts, Marion Petrie, and Tim D Spector, "Kin recognition signals in adult faces," *Vision research*, vol. 49, no. 1, pp. 38–43, 2009.

[19] Alexandra Alvergne, Ryo Oda, Charlotte Faurie, Akiko Matsumoto-Oda, Valérie Durand, and Michel Raymond, "Cross-cultural perceptions of facial resemblance between kin," *Journal of Vision*, vol. 9, no. 6, 2009.

[20] Daniel Brian Krupp, Lisa M DeBruine, and Benedict C Jones, "Cooperation and conflict in the light of 20 kin recognition systems," *The Oxford Handbook of Evolutionary Family Psychology*, p. 345, 2011.