Families in the Wild (FIW): Large-Scale Kinship Image Database and Benchmarks

Joseph P. Robinson, Ming Shao, Yue Wu, Yun Fu
Dept. of Electrical and Computer Engineering
Northeastern University
Boston, MA, USA
{jrobins1, mingshao, yuewu, yunfu}@ece.neu.edu

ABSTRACT

We present the largest kinship recognition dataset to date, Families in the Wild (FIW). Motivated by the lack of a single, unified dataset for kinship recognition, we aim to provide a dataset that captivates the interest of the research community. With only a small team, we were able to collect, organize, and label over 10,000 family photos of 1,000 families with our annotation tool designed to mark complex hierarchical relationships and local label information in a quick and efficient manner. We include several benchmarks for two image-based tasks, kinship verification and family recognition. For this, we incorporate several visual features and metric learning methods as baselines. Also, we demonstrate that a pre-trained Convolutional Neural Network (CNN) as an off-the-shelf feature extractor outperforms the other feature types. Then, results were further boosted by fine-tuning two deep CNNs on FIW data: (1) for kinship verification, a triplet loss function was learned on top of the network of pre-train weights; (2) for family recognition, a family-specific softmax classifier was added to the network.

1. INTRODUCTION

Automatic kinship recognition in visual media is essential for many real-world applications: *e.g.*, kinship verification [6, 8, 11, 26, 29, 30, 31, 32], automatic photo library management [21, 27], historic lineage and genealogical studies [2], social-media analysis [10], along with many of security applications addressing missing persons, human trafficking, crime scene investigations, and even our overall human sensing capabilities–ultimately, enhancing surveillance systems used in both real-time (*e.g.*, vBOLO [28] mission[1][2]) or offline (*e.g.*, searching for a subject in a large *gallery* [3, 23]). Thus, a gallery of imagery annotated with rich family information should lead to more powerful multimedia retrieval tools and complement many existing facial recognition systems (e.g., FBI's NGI[3]). However, even after several years (*i.e.*, since 2010 [8]) there are only a few vision systems capable of handling such tasks. Hence, kin-based technology in the visual domain has yet to truly evolve from research-to-reality.

[1]vBOLO: joint effort of two DHS Centers of Excellence, ALERT & VACCINE (http://www.northeastern.edu/alert)
[2]https://web-oup.s3-fips-us-gov-west-1.amazonaws.com/

MM '16, October 15-19, 2016, Amsterdam, Netherlands
© 2016 ACM. ISBN 978-1-4503-3603-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2964284.2967219

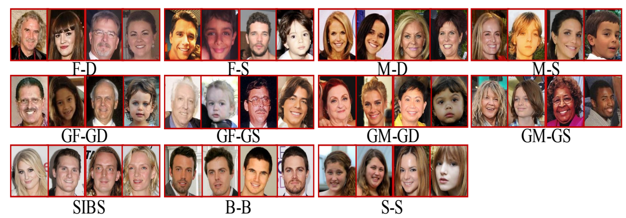

Figure 1: Sample faces chosen of 11 relationship types of FIW. *Parent-child*: (top row) Father-Daughter (F-D), Father-Son (F-S), Mother-Daughter (M-S) Mother-Son (M-S). *Grandparent-grandchild*: (middle row) same labeling convention as above. *Siblings*: (bottom row) Sister-Brother (SIBS), Brother-Brother (B-B), Sister-Sister (S-S).

We believe the reason that kinship recognition technology has not yet advanced to real-world applications is two-fold:

1. Current image datasets available for kinship tasks are not large enough to reflect the true data distributions of a family and their members.

2. Visual evidence for kin relationships are less discriminant than class types of other, more conventional machine vision problems (*e.g.*, facial recognition or object classification), as many hidden factors affect the similarities of facial appearances amongst family members.

To that end, we introduce a large-scale image dataset for kinship recognition called Families in the Wild (FIW). To the best of our knowledge, FIW is by far the **largest** and **most comprehensive** kinship dataset available in the vision and multimedia communities [see Table 1]. FIW includes 11,193 unconstrained family photos of 1,000 families, which is nearly 10x more than the next-to-largest, Family101 [7]. Also, it is from these 1,000 families that 418,060 image pairs for the 11 relationship types (see Figure 1 & Table 2).

Thus far, attempts at image-based kinship recognition have focused on facial features, as is the case for this work as well. Kinship recognition is typically conducted in one or more of the following modes: (1) kinship verification, which is a binary classification problem, *i.e.*, determine whether

[3]https://www.eff.org/2014/fbi-to-have-52M-face-photos

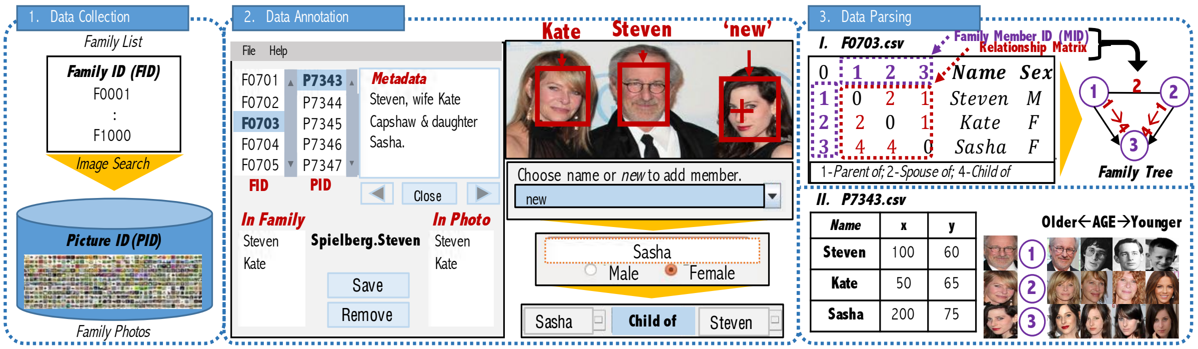

Figure 2: Method to construct FIW. ***Data Collection:*** **a list of candidate families (with an unique FID) and photos (with an unique PID) are collected.** ***Data Annotation:*** **a labeling tool optimized the process of marking the complex hierarchical nature of the 1,000 family trees of FIW.** ***Data Parsing:*** **post-processed the two sets of labels generated by the tool to partition data for kinship verification and family recognition.**

or not two or more persons are blood-related (kin or non kin); (2) family recognition, which is a multi-class classification problem where the aim is to determine the family of an individual in a photo; (3) kinship identification, which is a fine-grain categorization problem, as the aim is to determine the type of relationship shared between two or more people. We focus on the first and second modes in this paper, kinship verification and family recognition, which we briefly discuss next.

Kinship Verification. Previous efforts mainly focused on 4 *parent-child* pairs. As research in both psychology and computer vision revealed, different kin relations render different familiar features, and the 4 kin relations are usually modeled independently. Hence, it is important to have more kin relationship types accessible. Thus, FIW provides 11 pair-wise types (see Figure 1): 7 returning types (*i.e.*, *parent-child* and *siblings*), but in sample sizes scaling up 105x larger; and 4 types being offered to the research community for the first time (*i.e.*, *grandparent-grandchild*). Also, existing datasets contain, at most, only a couple of hundred pairs per category (*i.e.*, 1,000 in total). Such a insufficient amount leads to models overfitting the training data. Hence, existing models do not generalize well to new, unseen test data. However, FIW now makes 418,060 pairs available.

Family Recognition. A challenging task that grows more difficult with more families. This is because families contain large intra-class variations, often with overlaps between classes. Similar to conventional facial recognition, when the targets are unconstrained *faces in the wild* [12] (*i.e.*, pose, illumination, expression, and scene variant), the level of difficulty further increases, which is the same with kinship recognition. These are, unfortunately, challenges that need to be overcome. Thus, FIW poses realistic challenges needed to be surpassed before transitioning to real-world applications.

Contributions. We make three distinct contributions:

1. We introduce the largest visual kinship dataset to date, Families in the Wild (FIW).[4] FIW is complete with rich label information for 1,000 family trees, from which 11 kin relationship types were extracted in numbers that are orders of magnitude times larger than other datasets. This was made possible with an efficient annotation tool and procedure (Section 2).

2. We provide several benchmarks on FIW for both kinship verification and family recognition, including various low-level features, metric learning methods, and pre-trained Convolutional Neural Network (CNN) models. See Section 3 for experimental settings and results.

3. We fine-tune two CNNs: one with a triplet-loss layer on top, and the other with a softmax loss. Both yield a significant boost in performance over all other benchmarks for both tasks (Section 3.2).

[4]FIW will be available upon publication of this paper.

2. FAMILIES IN THE WILD

We now discuss the procedure followed to collect, organize, and label 11,193 family photos of 1,000 families with minimal manual labor. Then, we statistically compare FIW with other related datasets.

2.1 Building FIW

The goal for FIW was to collect approximately 10 photos for 1,000 families, each with at least 3 family members). We now summarize the method for achieving this in a three step process, which is visually depicted in Figure 2.

Step 1: Data Collection. A list of over 1,000 candidate families was made. To ensure diversity, we targeted groups of public figures worldwide by searching for <*ethnicity* OR *country*> AND <*occupation*> online (*e.g.*, *MLB* [Baseball] *Players*, *Brazilian Politicians*, *Chinese Actors*, *Denmark + Royal Family*). Family photos were then collected using various search engines (*e.g.*, Google, Bing, Yahoo) and social media outlets (*e.g.*, Pinterest) to widen the search space. Those with at least 3 family members and 8 family photos were added to FIW under an assigned Family ID (FID).

Step 2: Data Annotation. We developed a labeling tool to quickly annotate a large corpus of family photos. All photos for a given FID are labeled sequentially. Labeling is done by clicking a family member's face. Next, a face detector initializes a resizable box around the face. Faces unseen by the detector are discarded, as these are assumed to be poorly resolved. The tool then prompts for the name of the member via a drop-down menu. For starters, option new adds a member to a family under a unique member ID (MID), which prompts for the name, gender, and relationship types shared with others previously added to the current family (or FID). From there onward, labeling family members is just a matter of clicking.

Table 1: Comparison of FIW with related datasets.

Dataset	No. Family	No. People	No. Faces	Age Varies	Family Trees
CornellKin[8]	×	300	300	×	×
UBKinFace[20, 25]	×	400	600	✓	×
KFW-I[16]	×	1,066	1,066	×	×
KFW-II[16]	×	2,000	2,000	×	×
TSKinFace[18]	×	2,589	×	✓	✓
Family101[7]	101	607	14,816	✓	✓
FIW(Ours)	**1,000**	**10,676**	**30,725**	✓	✓

Table 2: Pair counts for FIW and related datasets.

	KFW-II [16]	Sibling Face [10]	Group Face [10]	Family 101[7]	FIW (Ours)
B-B	×	232	40	×	**86,000**
S-S	×	211	32	×	**86,000**
SIB	×	277	53	×	**75,000**
F-D	250	×	69	147	**45,000**
F-S	250	×	69	213	**43,000**
M-D	250	×	62	148	**44,000**
M-S	250	×	70	184	**37,000**
GF-GD	×	×	×	×	**410**
GF-GS	×	×	×	×	**350**
GM-GD	×	×	×	×	**550**
GM-GS	×	×	×	×	**750**
Total	1,000	720	395	607	**418,060**

Step 3: Dataset Parsing. Two sets of labels are generated in Step 2: (1) *image-level*, containing names and corresponding facial locations; (2) *family-level*, containing a relationship matrix that represents the entire family tree. Relationship matrices are then referenced to generate lists of member pairs for the 11 relationship types. Next, face detections are normalized and cropped with [14], then are stored according the previously assigned *FID*→ *MID*. Lastly, lists of image pairs are generated for both kinship verification and family recognition.

2.2 Database Statistics

Our FIW dataset far outdoes its predecessors in terms of quantity, quality, and purpose. FIW contains 11,193 family photos of 1,000 different families. There are about 10 images per family that include at least 3 and as many as 24 family members. We compare FIW to related datasets in Table 1 and 2. Clearly, FIW provides more families, identities, facial images, relationship types, and labeled pairs– the pair count of FIW is orders of magnitude bigger than the next-to-largest (*i.e.*, KFW-II).

3. EXPERIMENTS ON FIW

In this section, we first discuss visual features and related methods used to benchmark the FIW dataset. We then report and review all benchmark results. Finally, we discuss the two methods used to fine-tune the pre-trained CNN model: (1) training a triplet-loss for kinship verification and (2) learning a softmax classifier for family recognition. Both obtain top scores in the respective task.

3.1 Features and Related Methods

All features and methods covered here were used to benchmark FIW. First, we review handcrafted features, Scale Invariant Feature Transformation (SIFT) and Local Binary Patterns (LBP), which are both widely used in kinship verification [16] and facial recognition [19]. Next, we introduce VGG-Face, the pre-trained CNN model used here as an off-the-shelf feature extractor. Lastly, we review other related metric learning methods.

SIFT [**15**] features have been widely applied in object and face recognition. As done in [16], we resized all facial images to 64×64, and set the block size to 16×16 with a stride of 8. Thus, there were a total of 49 blocks for each image, yielding a feature vector of length $128 \times 49 = 6,272$D.

LBP [**1**] has been frequently used for texture analysis and face recognition, as it describes the appearance of an image in a small, local neighborhood around a pixel. Once again, we followed the feature settings of [16] by first resizing each facial image to 64×64, and then extracting LBP features from 16×16 non-overlapping blocks with a radius of 2 pixels and number of neighbors (*i.e.*, samples) set to 8. We then generated a 256D histogram from each, yielding a final feature vector of length $256 \times 16 = 4,096$D.

VGG-Face CNN [**17**] uses a "Very Deep" architecture with very small convolutional kernels (*i.e.*, 3×3) and convolutional stride (*i.e.*, 1 pixel). This model was pre-trained on over 2.6 million images of 2,622 celebrity identities. For this, each face image was resized to 224×224 and then fed-forward to the second-to-last fully-connected layer (*i.e.*, fc7) of the CNN model, producing a 4,096D feature vector.

Metric Learning methods are commonly employed in the visual domain, some designed specifically for kinship recognition [5, 16, 18] and some as generic metric learning methods [4]. We chose two representative methods from these two categories to report benchmarks on, Neighborhood Repulsed Metric Learning (NRML) [16] and Information Theoretic Metric Learning (ITML) [4].

3.2 Fine-Tuned CNN Model

Weights of deep networks are trained on larger amounts of generic source data, then fine-tuned on target data, which utilize a wider, more readily available source domain that resembles the target in either modality, view, or both [9]. Following this notion, and motivated by recent success with deep learning on faces [17, 22, 24], we fine-tuned the VGG-Face model, improving results for both kinship verification and family recognition.

Kinship Verification. In kinship verification, for each fold we select the families which have more than 10 images in the rest four folds. 90% images of each family are used to fine-tune the model and the rest for validation. An average of 8,295 images are selected. The experimental results of this part can be found in Table 3 and Figure 3. This is so far the best results on FIW. Specifically, we remove the last fully-connected layer which is used to identify 2,622 people and employed a triplet-loss [19] as the loss function. The second-to-last fully-connected layer was the only non-frozen layer of the original CNN model, with an initial learning rate of 10^{-5} that decreased by a factor of 10 every 700 iterations (out of 1,400). Batch size was 128 images, and other network settings were the same as the original VGG-Face model. Training was done on a single GTX Titan X with about 10GB GPU memory. Fine-tuning was done using the renown Caffe [13] framework.

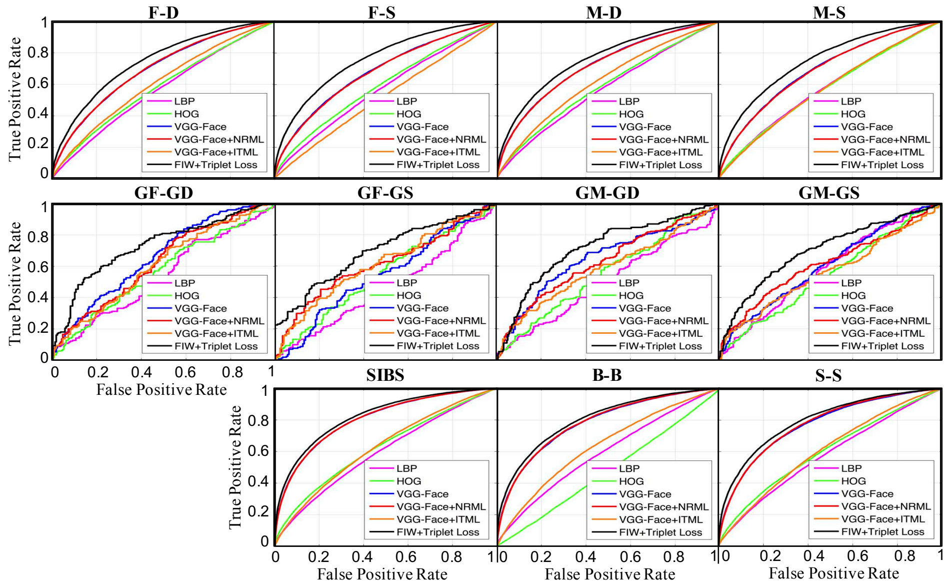

Figure 3: Relationship specific ROC curves depicting performance of each method.

Table 3: Verification accuracy scores (%) for 5-fold experiment on FIW. No family overlap between folds.

	F-D	F-S	M-D	M-S	SIBS	B-B	S-S	GF-GD	GF-GS	GM-GD	GM-GS	Avg.
SIFT	56.1	56.5	56.4	55.3	58.7	50.3	57.4	59.3	66.9	60.4	56.9	57.7±4.0
LBP	55.0	55.3	55.4	55.9	57.1	56.8	55.8	58.5	59.1	55.6	60.1	56.8±1.7
VGG-Face	64.4	63.4	66.2	64.0	73.2	71.5	70.8	64.4	68.6	66.2	63.5	66.9±3.5
Fine-Tuned CNN	**69.4**	**68.2**	**68.4**	**69.4**	**74.4**	**73.0**	**72.5**	**72.9**	**72.3**	**72.4**	**68.3**	**71.0±2.3**

Family Recognition. As visual kinship recognition is essentially related to facial features, we froze the weights in the lower levels of the VGG-Face network, and replaced with a new softmax layer to classify the 316 families. Other settings are the same as those used for kinship verification.

3.3 Experimental Settings

In this section, we provide benchmarks on FIW using all features and related methods mentioned in the previous section. Dimensionality of each feature is reduced to 100D using PCA. Experiments were done following a 5-fold cross-validation protocol. Each fold was of equal size and with no family overlap between folds.

Experiment 1: Kinship Verification. We randomly select an equal number of positive and negative pairs for each fold family. Cosine similarity is computed for each pair in the test fold. The average verification rate of all folds is reported in Table 3, showing that kinship verification is a challenging task. While some relation are relatively easy to recognize, *e.g.*, B-B, SIBS, S-S through SIFT, LBP, and VGG-Face features, results of other relations such as parent-child are still below 70.0%. Clearly, VGG-Face features are much better than hand-craft features. Notice grandparent-child pairs typically have higher accuracies than parent-child, which we believe is due to the differences in sample sizes. We also compare with the state-of-the-art metric learning methods, NRML and ITML [see Figure 3]. Showing improved scores for the low-level features, but still outperformed by the CNN model fine-tuned on FIW.

Table 4: Family Recognition accuracy scores (%) for 5-fold experiment on FIW (316 Families). No family overlap between folds.

	1	2	3	4	5	Avg.
VGG	9.6	14.5	11.6	12.7	13.1	12.3±1.8
Fine-tuned	**10.9**	**14.8**	**12.5**	**14.8**	**13.5**	**13.3±1.6**

Experiment 2: Family Recognition. We again follow the 5-fold cross-validation protocol with no family overlap. Families with 6 or more members, from which the 5 members with the most images were used. The results in Table 4 are from 316 families with 7,772 images. Folds were made up of one member for each family. Multi-class SVM was used to model VGG-Face features for each family (*i.e.*, one-vs-rest). We then improved the top-1 classification accuracy from 12.3±1.8 (%, VGG-Face) to 13.3±1.6 (%, our fine-tuned model).

4. DISCUSSION

We introduced a large-scale dataset of family photos captured in natural, unconstrained environments (*i.e.*, Families in the Wild). An annotation tool was designed to quickly generate rich label information for over 10,000 photos of 1,000 families. Emphasis was put on diversity (*i.e.*, families worldwide), data distribution (*i.e.*, at least 3 members and 8 photos per family), sample sizes (*i.e.*, multiple instances of each member, and at various ages), quality (*i.e.*, only faces seen by the detector), and quantity (*i.e.*, much more data and new relationship types).

There are many interesting directions for future work on our dataset. We currently only verify whether a pair of images is kin or non kin. However, also predicting the kin relationship type could lead to more value and practical usefulness. Thus, bringing us closer to doing fine-grain categorization on entire family trees. FIW will be an ongoing effort that will continually grow and evolve. See project page for downloads, updates, and to learn more about FIW.

Acknowledgements

This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 2013-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

We would also like to thank all members of SMILE Lab who helped with the process of collecting and annotating the FIW dataset.

5. REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(12):2037–2041, 2006.

[2] M. Almuashi, S. Z. M. Hashim, D. Mohamad, M. H. Alkawaz, and A. Ali. Automated kinship verification and identification through human facial images: a survey. *Multimedia Tools and Applications*, pages 1–43, 2015.

[3] N. Crosswhite, J. Byrne, O. M. Parkhi, C. Stauffer, Q. Cao, and A. Zisserman. Template adaptation for face verification and identification. *arXiv preprint arXiv:1603.03958*, 2016.

[4] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proc. International Conference on Machine Learning (ICML)*, pages 209–216. ACM, 2007.

[5] A. Dehghan, E. Ortiz, R. Villegas, and M. Shah. Who do i look like? determining parent-offspring resemblance via gated autoencoders. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1757–1764, 2014.

[6] H. Dibeklioglu, A. Salah, and T. Gevers. Like father, like son: Facial expression dynamics for kinship verification. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1497–1504, 2013.

[7] R. Fang, A. Gallagher, T. Chen, and A. Loui. Kinship classification by modeling facial feature heredity. In *20th IEEE International Conference on Image Processing (ICIP)*, pages 2983–2987. IEEE, 2013.

[8] R. Fang, K. D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *IEEE International Conference on Image Processing (ICIP)*, pages 1577–1580. IEEE, 2010.

[9] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013.

[10] Y. Guo, H. Dibeklioglu, and L. van der Maaten. Graph-based kinship recognition. In *International Conference on Pattern Recognition (ICPR)*, pages 4287–4292, 2014.

[11] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan. *Large Margin Multi-metric Learning for Face and Kinship Verification in the Wild*, pages 252–267. Springer International Publishing, Cham, 2015.

[12] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[14] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[16] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou. Neighborhood repulsed metric learning for kinship verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(2):331–345, 2014.

[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *Proc. British Machine Vision Conference (BMVC)*, 2015.

[18] X. Qin, X. Tan, and S. Chen. Tri-subject kinship verification: Understanding the core of A family. *CoRR*, abs/1501.02555, 2015.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.

[20] M. Shao, S. Xia, and Y. Fu. Genealogical face recognition based on ub kinface database. In *CVPR 2011 WORKSHOPS*, pages 60–65. IEEE, 2011.

[21] M. Shao, S. Xia, and Y. Fu. Identity and kinship relations in group pictures. In *Human-Centered Social Media Analytics*, pages 175–190. Springer, 2014.

[22] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1891–1898, 2014.

[23] D. Wang, C. Otto, and A. K. Jain. Face search at scale: 80 million gallery. *CoRR*, abs/1507.07242, 2015.

[24] M. Wang, Z. Li, X. Shu, J. Tang, et al. Deep kinship verification. In *IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2015.

[25] S. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *Proc. International Joint Conferences on Artificial Intelligence (IJCAI)*, volume 22, pages 2539–2544, 2011.

[26] S. Xia, M. Shao, and Y. Fu. Toward kinship verification using visual attributes. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 549–552. IEEE, 2012.

[27] S. Xia, M. Shao, J. Luo, and Y. Fu. Understanding kin relationships in a photo. *IEEE Transactions on Multimedia*, 14(4):1046–1056, 2012.

[28] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision (ECCV)*, pages 1–16. Springer, 2014.

[29] H. Yan, J. Lu, W. Deng, and X. Zhou. Discriminative multimetric learning for kinship verification. *IEEE Transactions on Information Forensics and Security*, 9(7):1169–1178, 2014.

[30] H. Yan, J. Lu, and X. Zhou. Prototype-based discriminative feature learning for kinship verification. *IEEE Transactions on Cybernetics*, 45(11):2535–2545, 2015.

[31] X. Zhou, J. Hu, J. Lu, Y. Shang, and Y. Guan. Kinship verification from facial images under uncontrolled conditions. In *Proc. ACM International Conference on Multimedia*, pages 953–956. ACM, 2011.

[32] X. Zhou, J. Lu, J. Hu, and Y. Shang. Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments. In *Proc. ACM International Conference on Multimedia*, pages 725–728. ACM, 2012.