

The FG 2015 Kinship Verification in the Wild Evaluation

Jiwen Lu, Junlin Hu, Venice Erin Liong, Xiuzhuang Zhou, Andrea Bottino,
Ihtesham Ul Islam, Tiago Figueiredo Vieira, Xiaoqian Qin, Xiaoyang Tan,
Songcan Chen, et al.

► **To cite this version:**

Jiwen Lu, Junlin Hu, Venice Erin Liong, Xiuzhuang Zhou, Andrea Bottino, et al.. The FG 2015 Kinship Verification in the Wild Evaluation. 11th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2015), May 2015, Ljubljana, Slovenia. hal-01158942

HAL Id: hal-01158942

<https://hal.archives-ouvertes.fr/hal-01158942>

Submitted on 2 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The FG 2015 Kinship Verification in the Wild Evaluation

Jiwen Lu¹, Junlin Hu², Venice Erin Liong¹, Xiuzhuang Zhou³, Andrea Bottino⁴,
Intesham Ul Islam⁴, Tiago Figueiredo Vieira⁴, Xiaoqian Qin⁵, Xiaoyang Tan⁵, Songcan Chen⁵,
Shahar Mahpod⁶, Yosi Keller⁶, Lilei Zheng⁷, Khalid Idrissi⁷, Christophe Garcia⁷,
Stefan Duffner⁷, Atilla Baskurt⁷, Modesto Castrillón-Santana⁸, Javier Lorenzo-Navarro⁸

¹Advanced Digital Sciences Center, Singapore

²School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

³College of Information Engineering, Capital Normal University, Beijing, China

⁴Politecnico di Torino, DAUIN, Torino, Italy

⁵Nanjing University of Aeronautics and Astronautics, Nanjing, China

⁶Faculty of Engineering, Bar Ilan University, Israel

⁷Université de Lyon, CNRS, INSA-Lyon, LIRIS, France

⁸SIANI-Universidad de Las Palmas de Gran Canaria, Spain

Email: jiwen.lu@adsc.com.sg

Abstract—The aim of the Kinship Verification in the Wild Evaluation (held in conjunction with the 2015 IEEE International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia) was to evaluate different kinship verification algorithms. For this task, two datasets were made available and three possible experimental protocols (unsupervised, image-restricted, and image-unrestricted) were designed. Five institutions submitted their results to the evaluation: (i) Politecnico di Torino, Italy; (ii) LIRIS-University of Lyon, France; (iii) Universidad de Las Palmas de Gran Canaria, Spain; (iv) Nanjing University of Aeronautics and Astronautics, China; and (v) Bar Ilan University, Israel. Most of the participants tackled the image-restricted challenge and experimental results demonstrated better kinship verification performance than the baseline methods provided by the organizers.

I. INTRODUCTION

This paper summarizes the results of the FG 2015 kinship verification in the wild evaluation. The objective of this evaluation is to determine whether there is a kin relation between a pair of given face images. The kinship is defined as a relationship between two persons who are biologically related with overlapping genes. Hence, there are four different types of kin relations investigated in this evaluation: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D). Face images are captured in uncontrolled environments in our datasets so that no restriction in terms of pose, lighting, background, expression, age, ethnicity, and partial occlusion on the images are used for training and testing.

Comparing with other face analysis such as face recognition [1]–[5], facial expression recognition [6], [7], and facial age estimation [8]–[11], kinship verification via faces is a relatively new problem in face analysis. Over the past five years, some kinship verification methods [12]–[31] have been

This study is partially supported by the research grant for the Human Sixth Sense Program (HSSP) at the Advanced Digital Sciences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore, and the research grant from the National Natural Science Foundation of China under grants 61373090.

TABLE I
EVALUATION SCHEDULE.

April 15, 2014	Evaluation announced
May 15, 2014	Datasets and features available
June 15, 2014	Registration deadline
November 5, 2014	Final results submitted
November 17, 2014	Summary paper submitted

proposed and most of them have shown that human face is an important cue to predict the kin relation between people. While some encouraging results have been obtained, most existing methods evaluated their systems and algorithms only with their own datasets and a standard protocol to compare different kinship verification methods is unarguably required.

In this work, we organize the kinship verification in the wild evaluation¹ by designing an open-set verification protocol to fairly evaluate and compare different kinship verification algorithms, which is held in conjunction with the IEEE International Conference on Automatic Face and Gesture Recognition 2015, Ljubljana, Slovenia. The schedule for the evaluation is shown in Table I. The shortened participant labels are used to simplify results summaries below. Our evaluation is designed with three different settings: unsupervised, image-restricted, and image-unrestricted. Five teams submitted their results to our evaluation and we summarize the details of the proposed methods in this paper. A summary of the participants submitting their final results is tabulated in Table II. We expect that the developed evaluation protocol in this work can advance the study of the kinship verification in the wild problem in the future.

II. DATASETS AND PROTOCOL

A. Data Sets

We provide two kinship face datasets: Kin Faces in the Wild I (KinFaceW-I) and Kin Faces in the Wild II

¹Available at: <http://www.kinfacew.com>

TABLE II
EVALUATION PARTICIPANTS.

Team	Country	Label
Politecnico di Torino	Italy	Polito
LIRIS-University of Lyon	France	LIRIS
Universidad de Las Palmas de Gran Canaria	Spain	ULPGC
Nanjing University of Aeronautics and Astronautics	China	NUAA
Bar Ilan University	Israel	BIU

(KinFaceW-II) [19], [26]. Face images in two datasets were collected from internet, including some public figures as well as their parents or children. Face images were captured under uncontrolled environments in two datasets with no restriction in terms of pose, lighting, background, expression, age, ethnicity, and partial occlusion. The difference of KinFaceW-I and KinFaceW-II is that face images with a kin relation were acquired from different photos in KinFaceW-I and the same photo in KinFaceW-II in most cases.

There are four kin relations in these two datasets: F-S, F-D, M-S, and M-D. In the KinFaceW-I dataset, there are 156, 134, 116, and 127 pairs of kinship images for these four relations. For the KinFaceW-II dataset, each relation contains 250 pairs of kinship images. For ease of use, we manually labeled the coordinates of the eyes position of each face image, and aligned and cropped facial region into 64×64 to remove background. Fig. 1 shows some cropped and aligned face images from KinFaceW-I and KinFaceW-II datasets, respectively.

B. Protocol

As a benchmark for comparison, we recommend using the pre-specified training/testing split, which was generated randomly and independently for 5-fold cross validation. Finally, the receiver operating characteristic (ROC) curve and the mean accuracy are used for performance evaluation. *Note that the training data and testing data are provided since beginning of this evaluation.*

Following the widely used unconstrained face verification benchmark [32], we designed the following three different settings to evaluate the performance of different kinship verification algorithms:

- Unsupervised setting: No labeled kin relation information is used.
- Image-restricted setting: Only the given kin relation information is used in the training splits.
- Image-unrestricted setting: The identity information of the person is available to potentially form additional negative pairs in the training splits.

III. BASELINE METHODS

In order to compare the performances of the different algorithms submitted, we also designed, for each evaluation setting, a specific baseline method. These methods are detailed in the following.

A. Features

For each face image, we extracted two different feature descriptors:

- **Local Binary Patterns (LBP)** [33]: we divided each image into 8×8 non-overlapping blocks, where the size of each block is 8×8 . We extracted a 59-dimensional uniform pattern LBP feature for each block and concatenated them to form a 3776-dimensional feature vector.
- **Histogram of Gradients (HOG)** [34]: we first divided each image into 16×16 non-overlapping blocks, where the size of each block is 4×4 , and then into 8×8 non-overlapping blocks, where the size of each block is 8×8 . Subsequently, we extracted a 9-dimensional HOG feature for each block and concatenated them to form a 2880-dimensional feature vector.

B. Models

We design the following three verification models for different settings in our evaluation:

- Unsupervised setting: Given a face pair, the cosine similarity of their LBP or HOG features is used to compute their similarity directly.
- Image-restricted setting: For each face image, we first apply PCA to project each LBP or HOG feature into a 500-dimensional feature vector and then side-information based linear discriminant analysis (SILD) [35] is employed to learn a distance metric. Specifically, the positive pairs and negative pairs in the training set were used to estimate the within-class and between-class variations of LDA. Finally, the cosine similarity of each test pair in the learned LDA space is computed.
- Image-unrestricted setting: For each face image, PCA is first used to project each LBP or HOG feature into a 500-dimensional feature vector and then neighborhood repulsed metric learning (NRML) [19], [26] is employed to learn a discriminative distance metric. Specifically, the label of training sample is used to seek the most similar intra-class neighbours to learn the distance metric. Finally, the cosine similarity of each test pair in the learned NRML space is computed.

IV. SUMMARY OF PARTICIPANTS' ALGORITHMS

This section briefly summarizes the algorithms presented by the five participating teams. Their results are compared with the baseline results presented by the organizers.

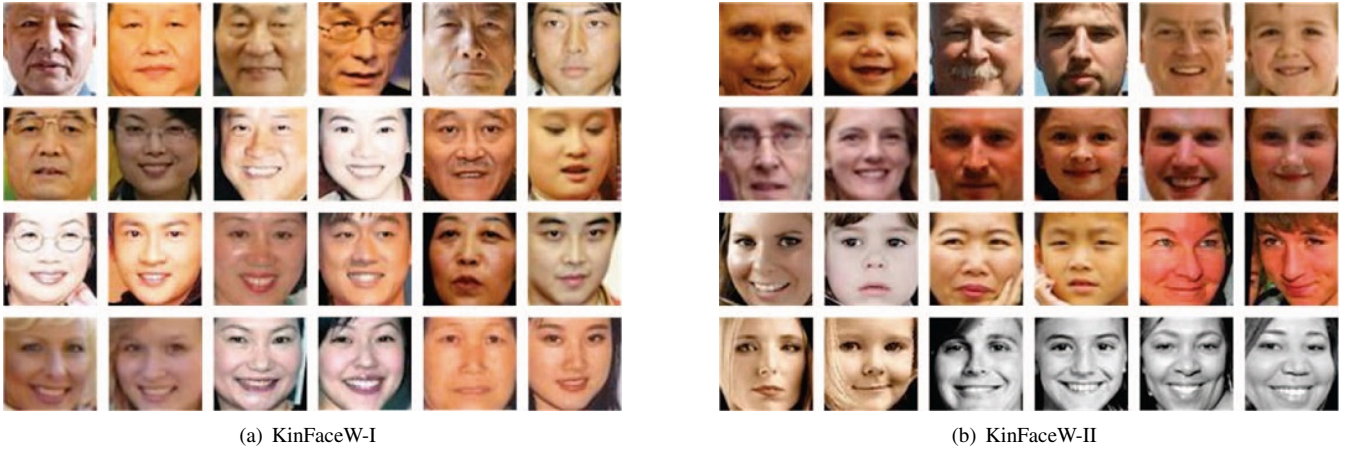


Fig. 1. Some cropped and aligned face images from the (a) KinFaceW-I and (b) KinFaceW-II datasets, respectively. From top to bottom are the F-S, F-D, M-S, and M-D kinship relations, and the two neighboring images in each row are with the kinship relation, respectively.

A. Politecnico di Torino

The Politecnico di Torino (Polito) group proposed a multi-perspective approach to the kinship verification under the image-restricted setting. Their method first combined different textural features and then selected the most relevant variables to provide an effective characterization of the samples. Lastly, the support vector machines (SVM) classifier is used for verification. Their method used three types of features to characterize each facial image:

- Local Phase Quantization (LPQ) [36] exploits the blur invariance property of the phase spectrum and encodes phase information. It is insensitive to centrally symmetric blur and is invariant to changes in the mean level of illumination and in illumination contrast. They applied a multi-scale LPQ representation to obtain a resulting vector of size 768.
- Three-Patch based LBPs (TBLBP) [37] encodes in a compact form of similarity between neighboring image patches. The basic idea behind patch-based approaches is to detect local texture properties by analyzing the cross-correlation between a central patch and its neighbouring patches. These properties are then encoded as a bit string. The image is encoded by dividing it into a set of non-overlapping blocks and concatenating the normalized histogram of code frequencies for each block. The final vector has size 3072.
- Weber Local Descriptor (WLD) [38] states that the difference in a stimulus can be noticed only if the ratio between this difference and the original stimulus exceeds a certain threshold derived from the Weber's law. WLD is based on two components computed on each pixel: the differential excitation and the orientation of the pixel gradient. Those components are extracted with a multiresolution analysis and then encoded into a histogram containing 2880 elements.

To improve classification accuracy, they applied the following three step feature selection (FS) process: 1) the features are ranked with the minimum-Redundancy-Maximum-

Relevance (mRMR) algorithm; 2) the size of the optimal mRMR set is heuristically found; and 3) the set of candidate features is further refined with a modified Sequential Forward Floating Selection algorithm. Finally, they performed classification using SVM with a radial basis kernel.

B. LIRIS-University of Lyon

This group applied the Triangular Similarity Metric Learning (TSML) [39] method on the problem of kinship verification by following the restricted setting on the two databases.

TSML uses a triplet $(\mathbf{x}_i, \mathbf{y}_i, s_i)$ to represent a pair of instances, where \mathbf{x}_i and \mathbf{y}_i are any two vectors, and $s_i = 1$ (resp. $s_i = -1$) means that the two vectors are similar (resp. dissimilar). It assumes a linear transformation function $f(\mathbf{x}, \mathbf{A})$ which projects the raw vectors into the target space: $\mathbf{a}_i = f(\mathbf{x}_i, \mathbf{A}) = \mathbf{A}\mathbf{x}_i$ and $\mathbf{b}_i = f(\mathbf{y}_i, \mathbf{A}) = \mathbf{A}\mathbf{y}_i$. The cost function of the TSML method is defined as:

$$J = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{2} \|\mathbf{a}_i\|^2 + \frac{1}{2} \|\mathbf{b}_i\|^2 - \|\mathbf{c}_i\| + 1 \right] + \frac{\lambda}{2} \|\mathbf{A} - \mathbf{A}_0\|^2, \quad (1)$$

where $\mathbf{c}_i = \mathbf{a}_i + s_i \mathbf{b}_i$ and $\frac{\lambda}{2} \|\mathbf{A} - \mathbf{A}_0\|^2$ is a regularization term to prevent overfitting. \mathbf{A}_0 is a pre-defined constant matrix, which is also the initialization for \mathbf{A} . The positive parameter λ adjusts the effects of the regularization term: the larger the parameter λ is, the closer \mathbf{A} is to \mathbf{A}_0 . Thus, this regularization guarantees the learnt optimal transformation function $f(\mathbf{x}, \mathbf{A}_*)$ to perform better than the initial function $f(\mathbf{x}, \mathbf{A}_0)$. The gradient of the above cost is:

$$\frac{\partial J}{\partial \mathbf{A}} = \frac{1}{n} \sum_{i=1}^n \left[\left(\mathbf{a}_i - \frac{\mathbf{c}_i}{\|\mathbf{c}_i\|} \right) x_i^T + \left(\mathbf{b}_i - \frac{s_i \mathbf{c}_i}{\|\mathbf{c}_i\|} \right) y_i^T \right] + \lambda (\mathbf{A} - \mathbf{A}_0). \quad (2)$$

Ideally, the optimal solution can be obtained at the zero gradient: $\mathbf{a}_i - \frac{\mathbf{c}_i}{\|\mathbf{c}_i\|} = 0$ and $\mathbf{b}_i - \frac{s_i \mathbf{c}_i}{\|\mathbf{c}_i\|} = 0$. In other words, the gradient function has set $\frac{\mathbf{c}_i}{\|\mathbf{c}_i\|}$ and $\frac{s_i \mathbf{c}_i}{\|\mathbf{c}_i\|}$ as targets for \mathbf{a}_i and \mathbf{b}_i , respectively. For a similar pair ($s_i = 1$), \mathbf{a}_i and \mathbf{b}_i are mapped to the same unit vector $\frac{\mathbf{c}_i}{\|\mathbf{c}_i\|}$; for a dissimilar pair ($s_i = -1$), \mathbf{a}_i and \mathbf{b}_i are mapped to opposite

unit vectors, $\frac{\mathbf{c}_i}{\|\mathbf{c}_i\|}$ and $-\frac{\mathbf{c}_i}{\|\mathbf{c}_i\|}$, respectively. In summary, minimizing the cost means to make similar vectors parallel and make dissimilar vectors opposite, which happens to be the objective of similarity metric learning.

They used four different face descriptors for each face image: LBP and HOG, the two features provided by our baseline method, plus Over-Complete LBP (OCLBP) [40] and Fisher Vector faces (FV) [41], two high-dimensional features with sizes of 22066 and 67584, respectively. Finally, each feature vector is transformed to new vectors with dimension 100 by whitened Principle Component Analysis (WPCA) for dimensionality reduction.

C. Universidad de Las Palmas de Gran Canaria

The ULPGC group adopted the image-restricted setting for kinship verification, where they made use of the multiple descriptors: Histogram of Oriented Gradients (HOG) [34], Local Binary Patterns (LBP) [33], Local Salient Patterns (LSP) [42], Local Directional Patterns (LDP) [43], Local Phase Quantization (LPQ) [36], Local Oriented Statistics Information Booster (LOSIB) [44]. By calculating descriptor for each image grids from 1×1 to 8×8 , they obtain a total of 15 descriptors. The total number of histograms per image is 120. Each parent-child pair, \mathbf{x}_i , is represented by the respective histogram similarities, $\mathbf{x}_i = \{S_{HOG}, S_{LBP}, \dots, S_{LOSIB}\}$. The similarity between two histograms \mathbf{h}^A and \mathbf{h}^B is defined as χ^2 distance:

$$\chi^2(\mathbf{h}^A, \mathbf{h}^B) = \sum_{c=1}^{n_{comps}} \frac{(h_c^A - h_c^B)^2}{(h_c^A + h_c^B)}. \quad (3)$$

For each kin-relation, PCA was performed for each fold. The number of principal components is chosen analyzing the accuracy achieved using a different number as input to train a SVM classifier.

D. Nanjing University of Aeronautics and Astronautics

The NUAA group contributed to this evaluation using image-restricted setting. Their method includes two parts: the base model, and the feature selection.

Model: The N training samples define a set $\{\mathbf{x}_i^p, \mathbf{x}_i^c, y_i\}_{i=1}^N$, where $\mathbf{x}_i^p, \mathbf{x}_i^c \in \mathbb{R}^d$ denotes the i -th sample of a parent and a child, d is the dimension of the sample, and $y_i \in \{+1, -1\}$ indicates whether the two individuals are related or not. The pairwise similarity between \mathbf{x}^p and \mathbf{x}^c is:

$$s^p(\mathbf{x}^p, \mathbf{x}^c) = (\mathbf{x}^p)^T \mathbf{W} \mathbf{x}^c. \quad (4)$$

And they formulate the following regularized logistic regression objective to optimize \mathbf{W} ,

$$\min_{\mathbf{W}, b} \sum_{i=1}^N \log(1 + \exp(-y_i(\mathbf{x}^p)^T \mathbf{W} \mathbf{x}^c)) + \lambda \|\mathbf{W}\|_*, \quad (5)$$

where b is the threshold, and $\|\mathbf{W}\|_*$ is the trace norm. This is a nonsmooth convex objective and one can use proximal methods to solve it, where at each step the singular values

of the standard gradient update are replaced by their soft-threshold versions.

Vote-based Feature Selection: To reduce of parameters of the proposed model, two steps are used for feature selection. In the first step, they evaluate the discriminative power of each feature of a parent regard to the given child. For this, they first partition an image into overlapping patches and extract a middle level 128-dimensional SIFT features from each patch, and a parent vector and a child vector is concatenated into a $2d$ -dimensional vector \mathbf{a} , and learn a weight vector \mathbf{u} with the same dimension using a sparse l_1 regularized logistic regression objective. This will give them a $2d$ -dimension vector \mathbf{u} with its first half (\mathbf{u}^p) and the second half (\mathbf{u}^c) respectively representing the importance of each feature of the parent and the child. Then, they use \mathbf{u}^p to vote the patches of parent face images and select those patches receiving high votes for face representation, and similar case to child images. Finally, linear SVM is utilized for final judgement.

E. Bar Ilan University

The BIU team proposed an asymmetric metric learning scheme with respect to kin similarity under both image-restricted and image-unrestricted settings. The gist of this approach is that kin verification is inherently asymmetric where one (father, mother) is a priori known to be significantly older than the other (son, daughter), resulting in significant appearance differences. Thus, they propose to learn *two* projections such that the distance is given by

$$d_{\mathbf{a}}^2(\phi_i^o, \phi_j^y) = \|\mathbf{W}_o \phi_i^o - \mathbf{W}_y \phi_j^y\|_2^2, \quad (6)$$

where \mathbf{W}_o is the projection matrix applied to *old* samples (father, mother), and \mathbf{W}_y is the matrix applied to *young* samples (son, daughter). $\phi_i^o \in \mathbb{R}^d$ and $\phi_j^y \in \mathbb{R}^d$ are the features based representation of the faces of old and young subjects, respectively, where it is a priori known which images depict either of these two classes. Training and test samples are given in positive and negative pairs, such that $\{\phi_i^o, \phi_i^y\}$ is a positive pair corresponding to the i^{th} family, while $\{\phi_i^o, \phi_j^y\}$, $i \neq j$, is a negative pair.

To learn metric pair $\mathbf{W} = \{\mathbf{W}_o, \mathbf{W}_y\}$, they formulate the objective function by a margin maximization scheme as:

$$\arg \min_{\mathbf{W}, b} \sum_{i,j} \max[1 - y_{ij}(b - d_{\mathbf{a}}^2(\phi_i^o, \phi_j^y)), 0], \quad (7)$$

where $y_{ij} = 1$ if the subjects in i and j are kins, and $y_{ij} = -1$ otherwise; b is set as the mid point between the average distances of related and unrelated subjects. They solved this problem by a stochastic subgradient scheme.

Having obtained the optimal metric \mathbf{W} , the SVM classifier with the RBF kernel using the distance $d_{\mathbf{a}}^2$ between samples is used to verify kin relationship. The features used by this method are the baseline HOG and LBP features.

V. RESULTS

Figs. 2-7 shows the ROC curves of different methods on different settings on our datasets, and Tables III – VIII

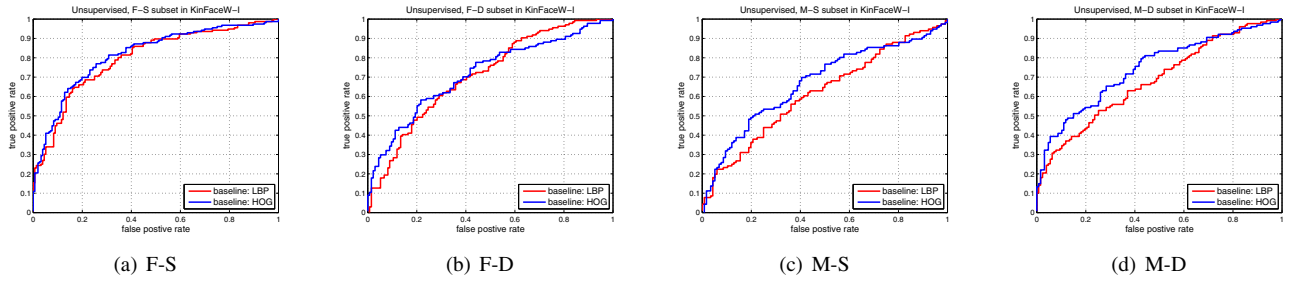


Fig. 2. ROC curves of the baseline method with the unsupervised setting on KinFaceW-I.

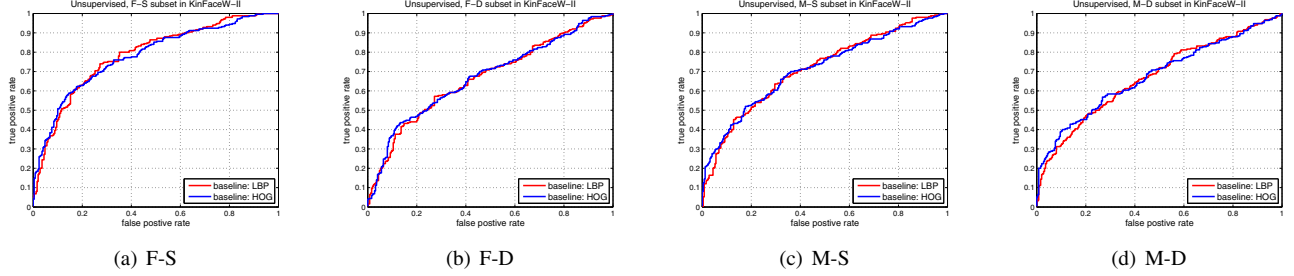


Fig. 3. ROC curves of the baseline method with the unsupervised setting on KinFaceW-II.

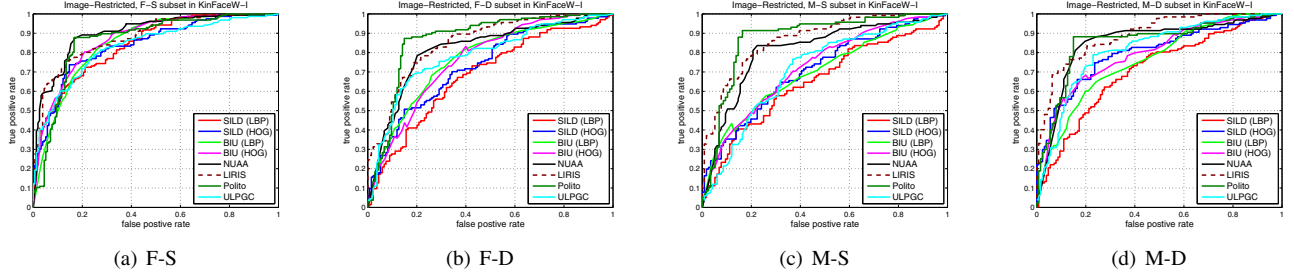


Fig. 4. ROC curves of different methods under the image-restricted setting on KinFaceW-I.

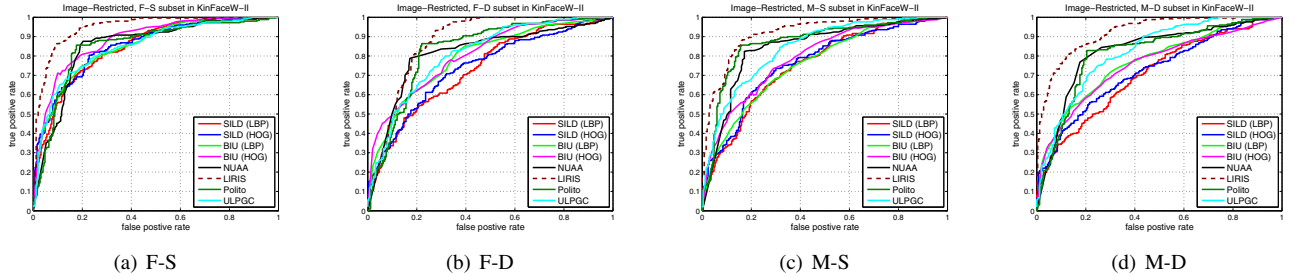


Fig. 5. ROC curves of different methods under the image-restricted setting on KinFaceW-II.

TABLE III

THE MEAN ACCURACY (%) UNDER UNSUPERVISED SETTING ON THE KINFACEW-I DATASET.

Label	F-S	F-D	M-S	M-D	Mean
SILD (LBP)	76.92	69.05	64.24	66.59	69.20
SILD (HOG)	79.53	70.93	67.66	72.84	72.74

TABLE IV

THE MEAN ACCURACY (%) UNDER UNSUPERVISED SETTING ON THE KINFACEW-II DATASET.

Label	F-S	F-D	M-S	M-D	Mean
SILD (LBP)	75.40	66.60	70.60	66.00	69.65
SILD (HOG)	74.20	66.60	70.60	67.00	69.60

list the mean accuracy of the three different settings. From the experimental results, we can see that 1) Only one team, namely BIU evaluates their method under image-unrestricted setting and achieves better verification results

than baseline NRML on all four subsets of two datasets; 2) All five groups contribute to the challenge under the image-restricted setting, team Polito achieves the first place (86.30%) on the KinFaceW-I and the second place (83.10%)

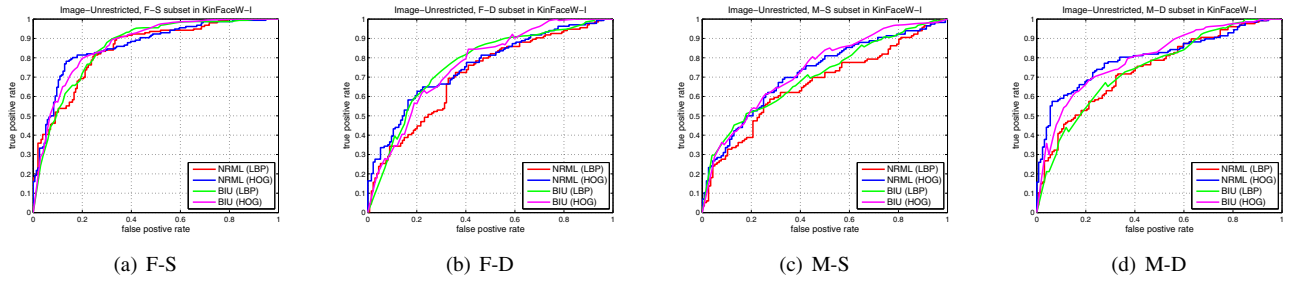


Fig. 6. ROC curves of different methods under the image-unrestricted setting on KinFaceW-I.

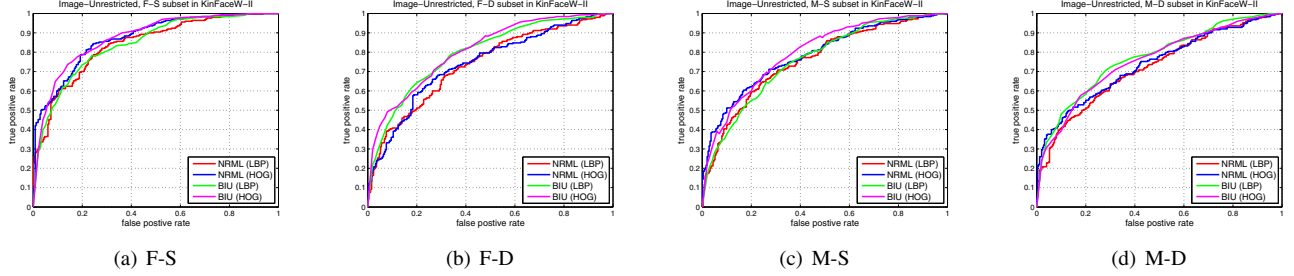


Fig. 7. ROC curves of different methods under the image-unrestricted setting on KinFaceW-II.

TABLE V

THE MEAN ACCURACY (%) UNDER IMAGE-RESTRICTED SETTING ON THE KINFACEW-I DATASET.

Label	F-S	F-D	M-S	M-D	Mean
Polito	85.30	85.80	87.50	86.70	86.30
LIRIS	83.04	80.63	82.30	84.98	82.74
ULPGC	71.25	70.85	58.52	80.89	70.01
NAAA	86.25	80.64	81.03	83.93	82.96
BIU	86.90	76.48	73.89	79.75	79.25
SILD (LBP)	78.22	69.40	66.81	70.10	71.13
SILD (HOG)	80.46	72.39	69.82	77.10	74.94

TABLE VI

THE MEAN ACCURACY (%) UNDER IMAGE-RESTRICTED SETTING ON THE KINFACEW-II DATASET.

Label	F-S	F-D	M-S	M-D	Mean
Polito	84.00	82.20	84.80	81.20	83.10
LIRIS	89.40	83.60	86.20	85.00	86.05
ULPGC	85.40	75.80	75.60	81.60	80.00
NAAA	84.40	81.60	82.80	81.60	82.50
BIU	87.51	80.82	79.78	75.63	80.94
SILD (LBP)	78.20	70.00	71.20	67.80	71.80
SILD (HOG)	79.60	71.60	73.20	69.60	73.50

TABLE VII

THE MEAN ACCURACY (%) UNDER IMAGE-UNRESTRICTED SETTING ON THE KINFACEW-I DATASET.

Label	F-S	F-D	M-S	M-D	Mean
BIU (LBP)	85.51	76.54	69.93	74.36	76.59
BIU (HOG)	86.90	76.48	70.62	79.75	78.44
NRML (LBP)	81.43	69.76	67.23	72.87	72.82
NRML (HOG)	83.68	74.64	71.56	79.96	77.46

TABLE VIII

THE MEAN ACCURACY (%) UNDER IMAGE-UNRESTRICTED SETTING ON THE KINFACEW-II DATASET.

Label	F-S	F-D	M-S	M-D	Mean
BIU (LBP)	84.24	79.45	75.98	77.04	79.18
BIU (HOG)	87.51	80.82	79.78	75.63	80.94
NRML (LBP)	79.20	71.60	72.20	68.40	72.85
NRML (HOG)	80.80	72.80	74.80	70.40	74.70

- While encouraging performance has been obtained in our evaluation, there is still much space to improve for kinship verification in the wild, especially compared with the current state-of-the-arts in face verification.

VI. CONCLUSION AND FUTURE WORK

on the KinFaceW-II in term of mean verification rate, while group LIRIS obtains the best result on the KinFaceW-II dataset for all four subsets. In general, all teams show better kinship verification performance than the baseline methods in this evaluation. In addition, we can make the following observations from the above results:

- The unsupervised setting is the most challenging one and the image-unrestricted setting is the easiest. This is because the image-unrestricted setting can utilize more discriminative information than the other two settings.

This paper describes the kinship verification in the wild evaluation which is held in conjunction with the 2015 IEEE International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia. This evaluation has been a great community effort because most existing kinship verification methods were only evaluated with their own datasets and a standard protocol to compare different kinship verification methods is required. In this work, we have established such a benchmark for kinship verification via face images, which will allow researchers in this field to further

investigate this problem. To ensure the benchmark available in the future, the organizers are keeping the datasets available through their online repository.

REFERENCES

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [3] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative multimifold analysis for face recognition from a single training sample per person," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 39–51, 2013.
- [4] J. Hu, J. Lu, and Y. Tan, "Discriminative deep metric learning for face verification in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1875–1882, 2014.
- [5] J. Lu, G. Wang, W. Deng, and K. Jia, "Reconstruction-based metric learning for unconstrained face verification," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 79–89, 2015.
- [6] B. Fasel and J. Luetttin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.
- [7] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 160–187, 2003.
- [8] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.
- [9] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.
- [10] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.
- [11] J. Lu and Y. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 2, pp. 249–258, 2013.
- [12] R. Fang, K. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," in *IEEE International Conference on Image Processing*, pp. 1577–1580, 2010.
- [13] X. Zhou, J. Hu, J. Lu, Y. Shang, and Y. Guan, "Kinship verification from facial images under uncontrolled conditions," in *ACM International Conference on Multimedia*, pp. 953–956, 2011.
- [14] S. Xia, M. Shao, and Y. Fu, "Kinship verification through transfer learning," in *International Joint Conference on Artificial Intelligence*, pp. 2539–2544, 2011.
- [15] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding kin relationships in a photo," *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 1046–1056, 2012.
- [16] G. Somanath and C. Kambhamettu, "Can faces verify blood-relationships?," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pp. 105–112, 2012.
- [17] S. Xia, M. Shao, and Y. Fu, "Toward kinship verification using visual attributes," in *International Conference on Pattern Recognition*, pp. 549–552, 2012.
- [18] G. Guo and X. Wang, "Kinship measurement on salient facial features," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2322–2325, 2012.
- [19] J. Lu, J. Hu, X. Zhou, Y. Shang, Y.-P. Tan, and G. Wang, "Neighborhood repulsed metric learning for kinship verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2594–2601, 2012.
- [20] N. Kohli, R. Singh, and M. Vatsa, "Self-similarity representation of weber faces for kinship classification," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pp. 245–250, 2012.
- [21] X. Zhou, J. Lu, J. Hu, and Y. Shang, "Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments," in *ACM International Conference on Multimedia*, pp. 725–728, 2012.
- [22] R. Fang, A. C. Gallagher, T. Chen, and A. Loui, "Kinship classification by modeling facial feature heredity," in *IEEE International Conference on Image Processing*, pp. 2983–2987, 2013.
- [23] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Like father, like son: Facial expression dynamics for kinship verification," in *IEEE International Conference on Computer Vision*, pp. 1497–1504, 2013.
- [24] T. F. Vieira, A. Bottino, and I. U. Islam, "Automatic verification of parent-child pairs from face images," in *The 18th Iberoamerican Congress on Pattern Recognition, CIARP 2013*, pp. 326–333, 2013.
- [25] T. F. Vieira, A. Bottino, A. Laurentini, and M. D. Simone, "Detecting siblings in image pairs," *The Visual Computer*, vol. 30, no. 12, pp. 1333–1345, 2014.
- [26] 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*, vol. 34, no. 2, pp. 331–345, 2014.
- [27] H. Yan, J. Lu, W. Deng, and X. Zhou, "Discriminative multimetric learning for kinship verification," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1169–1178, 2014.
- [28] Y. Guo, H. Dibeklioglu, and L. van der Maaten, "Graph-based kinship recognition," in *International Conference on Pattern Recognition*, pp. 4287–4292, 2014.
- [29] J. Lu, J. Hu, X. Zhou, J. Zhou, M. Castrillon-Santana, J. Lorenzo-Navarro, L. Kou, Y. Shang, A. Bottino, and T. F. Vieira, "Kinship verification in the wild: The first kinship verification competition," in *IEEE International Joint Conference on Biometrics*, pp. 1–6, 2014.
- [30] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan, "Large margin multi-metric learning for face and kinship verification in the wild," in *Asian Conference on Computer Vision*, 2014.
- [31] H. Yan, J. Lu, and X. Zhou, "Prototype-based discriminative feature learning for kinship verification," *IEEE Transactions on Cybernetics*, 2014.
- [32] 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," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.
- [33] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [34] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.
- [35] M. Kan, S. Shan, D. Xu, and X. Chen, "Side-information based linear discriminant analysis for face recognition," in *British Machine Vision Conference*, pp. 1–12, 2011.
- [36] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä, "Recognition of blurred faces using local phase quantization," in *International Conference on Pattern Recognition*, pp. 1–4, 2008.
- [37] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Real-Life Images workshop at the European Conference on Computer Vision*, October 2008.
- [38] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.
- [39] L. Zheng, K. Idrissi, C. Garcia, S. Duffner, and A. Baskurt, "Triangular similarity metric learning for face verification," in *11th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2015)*, 2015.
- [40] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *IEEE International Conference on Computer Vision*, pp. 1960–1967, 2013.
- [41] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *British Machine Vision Conference*, 2013.
- [42] Z. Chai, Z. Sun, T. Tan, and H. M. Vazquez, "Local salient patterns - A novel local descriptor for face recognition," in *International Conference on Biometrics*, pp. 1–6, 2013.
- [43] T. Jabit, M. Kabir, and O. Chae, "Local directional pattern (ldp) for face recognition," in *2010 Digest of Technical Papers International Conference on Consumer Electronics*, pp. 329–330, 2010.
- [44] O. García-Olalla, E. Alegre, L. Fernández-Robles, and V. González-Castro, "Local oriented statistics information booster (LOSIB) for texture classification," in *International Conference on Pattern Recognition*, pp. 1114–1119, 2014.