Tri-Subject Kinship Verification: Understanding the Core of A Family

Xiaoqian Qin, Xiaoyang Tan, and Songcan Chen

Abstract—One major challenge in computer vision is to go beyond the modeling of individual objects and to investigate the bi- (one-versus-one) or tri- (one-versus-two) relationship among multiple visual entities, answering such questions as whether a child in a photo belongs to the given parents. The child-parents relationship plays a core role in a family, and understanding such kin relationship would have a fundamental impact on the behavior of an artificial intelligent agent working in the human world. In this work, we tackle the problem of one-versus-two (tri-subject) kinship verification and our contributions are threefold: 1) a novel relative symmetric bilinear model (RSBM) is introduced to model the similarity between the child and the parents, by incorporating the prior knowledge that a child may resemble one particular parent more than the other; 2) a spatially voted method for feature selection, which jointly selects the most discriminative features for the child-parents pair, while taking local spatial information into account; and 3) a large-scale tri-subject kinship database characterized by over 1,000 child-parents families. Extensive experiments on KinFaceW, Family101, and our newly released kinship database show that the proposed method outperforms several previous state of the art methods, while could also be used to significantly boost the performance of one-versus-one kinship verification when the information about both parents are available.

Index Terms—Feature selection, kinship verification, tri-subject relationship.

I. INTRODUCTION

KINSHIP verification from facial images is an emerging problem in computer vision. From an aspect of face recognition, kinship provides us with a valuable and operational opportunity to construct useful relationship between persons based on their visual signals, thus deepening our understanding on their semantics. Applications of kin relationships include face image retrieval [1]–[3]/annotation [4], [5]/organization, increasing face recognition rates [6], [7], social media analysis [8], [9], finding of missing children, children adoptions [10], and so on.

Besides its wide applications, kinship learning is also motivated by the long-term goal of computer vision to go beyond the understanding of a single visual entity (e.g., "whose face is this?") and to investigate the bi- or tri- relationship among multiple visual entities, e.g., answering such questions as whether a child in a photo belongs to given parents. Actually, recent research has demonstrated that computer vision algorithms have been able to understand individual face image fairly well - the best result on the challenging LFW (labeled face in the wild) face verification database has reached an accuracy as high as 99.15% [11] - even better than what can be done by a human being. However, extending those techniques to characterise the complex relationship among multiple entities is not trivial. One major reason is due to the fact the appearance gap encountered in a kinship problem is much larger than that in a conventional face recognition setting (e.g, given two face images with different sex and different ages, verify whether those two subjects are father and daughter).

In this sense, kinship learning is a step towards such a trend to capture mutual information among different visual entities, particularly multiple face images. Most of current researches [12]–[17], however, mainly focus on the kinship involving only two subjects (one-versus-one) such as father-son or mother-daughter, while in practice, kin relationship involving more subjects are desirable, for example, in the problem of finding missing children, usually we have the photos of both parents, and there is no reason preventing us from using images of both parents at the same time for more effective kinship verification. As another application scenario of law enforcement, it would be beneficial to match the image of a criminal suspect with those of his/her parents to improve the performance of suspect searching. Motivated by this, [18] assembled a family database containing 45 families with an average of 120 near frontal facial samples per family. Fang *et al.* [19] collected the Family101 kinship dataset, containing 14,816 face images from 206 nuclear families. Both [18] and [19] ask questions concerning more general family membership (one-versus-multi) beyond father and son.

In this paper we focus on the problem of tri-subject (one-versus-two) kinship learning (i.e., son-parents and daughter-parents). This is an important special case of the more ambitious one-versus-multi verification and is largely overlooked in literatures. The child-parents is the core and the most basic unit formed in a family and understanding such kind of kin relationship would have fundamental impact on the behavior of an artificial intelligent agent working in a human world. Furthermore,

X. Qin is with the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, and also with Huaiyin Normal University, Huaiyin 223300, China (e-mail: qinxiaoqian@hytc.edu.cn).

X. Tan and S. Chen are with the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: x.tan@nuaa.edu.cn; s.chen@nuaa.edu.cn).

compared to the problem of one-versus-multi kinship verification, the one-vs-two verification is a more convenient and more practical choice - not only because its scope is more controllable, but also because the problem by itself is easier to define since otherwise it could be difficult to determine kinship relations in a big family genetically and without ambiguity, especially for those people among whom the kinship ties are weak.

To address this problem of tri-subject kinship verification, the key idea of our method is to fully exploit the dependence structure between child and parents in a few aspects: similarity measure, feature selection and classifier design. This is based on the observation that compared to the case with only one image from one of the parents, images from both parents could provide richer information about the kinship relation regards to a child, due to the genetic overlapping between a pair of parents and their child. To this end, our contributions are three folds. First, we use a bilinear function to model the similarity between the parents and the child, with the dependence between them captured by a covariance-like matrix learnt from the data. To make this more robust, we introduce a novel relative bilinear similarity model which effectively incorporates the prior knowledge [20] that children may resemble a particular parent more than another.

Second, we propose a spatially voted method for feature selection, which jointly selects the most discriminative features for the child-parents pair, while taking local spatial information into account. Compared to traditional group-based feature selection methods such as group lasso, we essentially allow the features in a whole image to compete with each other and then select the group in which higher portion of individual features in the corresponding local region win. By contrast, in group lasso, features are teamed together beforehand and have to compete with others as a group. Our method is more flexible than the latter in the sense that it permits fine-grained control over the contribution of each feature to the establishment of one-vs-two kin relationships.

Finally, we release a new face database specific to the tri-subject kinship problem, characterized by over 1,000 child-parents groups. State-of-the-art results are achieved using our method. Interestingly, our experimental results also show that the accuracy of one-vs-one bi-kinship verification benefits a lot by reformulating it as a specific case of one-vs-two tri-kinship verification when the information about a second parent is available.

This journal paper builds on the earlier conference work [21]. In what follows we briefly review some of the related work in Section II, and detail our proposed method in Section III. Our new kinship database is described in Section IV and experimental results are given in Section V. We conclude this paper in Section VI.

II. RELATED WORK

The aim of bi-subject (one-versus-one) kinship verification through computer vision is predicting whether a given pair of images has kin relation. The research in the field of human visual signal processing [22], [23] has provided strong evidence that facial appearance is a useful cue for genetic similarity, since children look more similar to their parents than other adults of the same gender. To find such distinguishable cues from facial appearance, in an early attempt, Fang *et al.* [12] used various features including the skin, hair and eye color, facial structure measures and local/holistic texture.

Later, researchers evaluated various types of feature descriptors for kinship verification. In [10], the DAISY descriptors are adopted to facilitate local facial patches matching for eyes, mouth and nose with spatial Gaussian kernels. In [24], a spatial pyramid learning-based feature descriptor is utilized to represent kinship faces. In [25], a gated autoencoder method is used to encode the resemblance between a parent and a child, which is trained through minimizing the reconstruction error given a set of randomly sampled local patches. In [7], dense stereo matching is used to determine kinship similarity. Other feature sets for kinship verification include Gradient Orientation Pyramid (GGOP) [8], Self Similarity Representation (SSR) [26] and prototype-based discriminative feature learning (PDFL) method [27]. Since semantic-related feature sets such as attributes usually show more tolerance to appearance changes, they are naturally used for kinship verification [14]. Based on the idea that people look more like their parents when they smile, [16] proposes to describe facial dynamics and spatio-temporal appearance over smile expression and uses these to improve the kinship verification rate.

In [15] the authors show that combining several types of middle-level features is useful. For that purpose, they introduced a multiview neighborhood repulsed metric learning method (MNRML) by learning a distance metric under which the samples with a kinship relation are pulled close and those without a kinship relation are pushed away. [17] and [28] extract multiple features to characterize face images and maximize the correlation of different features to exploit complementary information for kinship verification. Another way to reduce the appearance similarity gap is to use intermediate samples which bridge the two sides with large divergence. In [13], [29], [30] and [31], such a bridge is constructed by facial images of parents at the similar ages of their children. However, it is not easy to collect such an image set in practice.

While most of the above works focus on the bi-subject (one-versus-one) kinship verification, [18] and [19] deal with the one-versus-multi kinship relation. Particularly Ghahramani *et al.* [18] addresses the problem of family verification, i.e., predicting whether a query face image has kin relation with multiple family members, by fusing similarity of each member's facial image segments. Fang *et al.* [19] tackle the more general family membership classification, i.e., given a query face image, asking which family it belongs to, and they do this with a minimum sparse reconstruction method. Despite the partial success of these methods, we argue that in general it is difficult to establish the relationship between a subject and some members of his/her family through the face appearance if the kinship ties between them is weak.[1] Instead we focus on the verification of the most basic unit that forms a family, that is, the child-parents (one-versus-two) relationship. We call this tri-subject kinship verification. The methods developed here can be potentially extended to handle more complex relationship by treating a family tree as an ensemble of tri-relationships.

---

[1]For example, it makes no sense to reconstruct a man's face image using his father-in-law's.
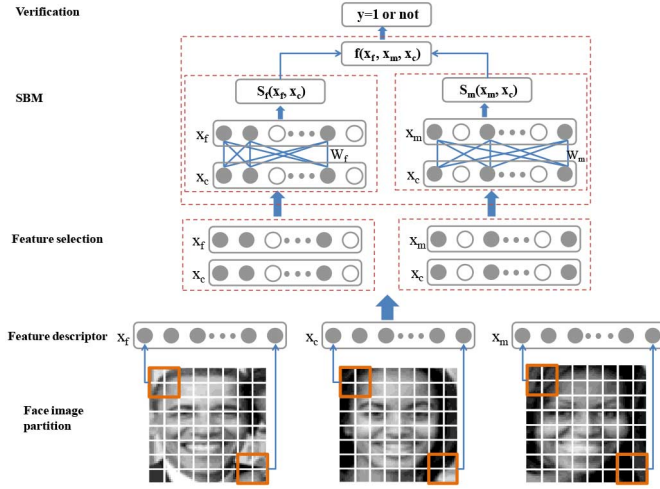
Fig. 1. Overall architecture of the proposed method.

III. TRI-SUBJECT KINSHIP VERIFICATION

In this section, we present our method for tri-subject kinship verification. Assume that we are given a set of N training samples $\{(x_{fi}, x_{mi}, x_{ci}, y_i)\}_{i=1}^N$, where $x_{fi}, x_{mi}, x_{ci} \in R^d$ respectively denotes the i-th sample of a father, a mother and a child, d is the dimension of the feature representation of a sample, and $y_i \in \{+1, -1\}$ indicates whether this child has a valid kinship relation with the corresponding two parents. Here by kinship relation we mean a very close family type relation, that is, the child is produced by the two parents.

Our goal is to learn a function $f : (x_f, x_m, x_c) \rightarrow \{+1, -1\}$ from the training data to check whether such a kinship could be established for three previously never seen images (x_f, x_m, x_c) of a couple and a child. For simplicity we assume that the gender of both parents images (x_f, x_m) are known and that they indeed genetically produce some children, but we do not know whether x_c is one of them. We also assume that the gender of the test image x_c of the child is known.

A. Two Bilinear Models for Kinship Verification

The overall architecture of the proposed method is shown in Fig. 1, which can be roughly divided into three stages. Particularly, in the first stage, we partition an image into overlapping patches and extract a middle-level feature descriptor (e.g., 128-dimensional SIFT features) from each patch (location), which are then concatenated into a feature vector as the input to the next stage. In the second stage we use a spatially voted feature selection method to select the most discriminative local facial patches to improve the robustness. Finally in the third stage, we learn the similarity between parents and child using bilinear models, based on which the final kinship verification is made.

In this work, we explore two ways to encode the similarity between parents and a child. The first one is to decompose the triples of (x_f, x_m, x_c) into two pairs (x_f, x_c) and (x_m, x_c), and the pairwise similarity between them is respectively

$$s_f(x_f, x_c) = (x_f)^T W_f x_c \equiv \langle x_f, x_c \rangle_{W_f}$$
$$s_m(x_m, x_c) = (x_m)^T W_m x_c \equiv \langle x_m, x_c \rangle_{W_m} \quad (1)$$

where $\langle a, b \rangle_W \equiv a^T W b$, and the transformation matrix W_f, W_m essentially encodes the "covariance" relationship between a parent and a child, to be learnt from the training data.

Since both W_f, W_m are $d \times d$ matrix and the similarity function is a bilinear function, we call this Symmetric Bilinear Model (SBM). The bilinear model has many advantages compared to the simple Euclidean-based model: 1) it is a natural choice to model the similarity between two subjects; and 2) it is also a much richer model than a traditional linear model–actually the bilinear model is related to the Mahalanobis distance (especially when the energy of each feature vector is fixed) and hence it can effectively capture the correlation between any two feature variables. However, the bilinear model is different from the Mahalanobis distance in that its parameter matrix W is not necessarily a positive definite matrix, which not only indicates that it could be more flexible than a traditional metric learning-based method, but also means that what a bilinear model learns is not a metric but a classifier. But this is exactly what we need–a model to predict directly whether a given pair of subjects has some kind of kinship, rather than the metric between them.

We further denote the probability that a child x_c belongs to a pair of parents (x_f, x_m) as $P(y = 1|x_f, x_m, x_c)$, and it is linked to our verification function $f(x_f, x_m, x_c)$ through a sigmoid function, i.e.

$$p(y = 1|x_f, x_m, x_c) = \sigma(f(x_f, x_m, x_c)) \quad (2)$$

where sigmoid function σ is defined to be $\sigma(x) = \frac{1}{1+e^{-x}}$. The verification function $f(x_f, x_m, x_c)$ is modeled as the linear combination of two pieces of evidence, i.e., the similarity of x_c to x_m and x_f, respectively

$$f(x_f, x_m, x_c) = \beta_1 s_f(x_f, x_c) + \beta_2 s_m(x_m, x_c) + b \quad (3)$$

where the combination coefficients β_1 and β_2 are two scalars and b is the similarity threshold term. To learn these parameters, we maximize the conditional likelihood defined by (2) by plugging (1) into it, with L_2 regularization added. How to learn the pairwise similarity (1) will be detailed in the next section.

Alternatively, one can treat the parents and the child as samples from two domains. Let us denote the parents domain as $\mathcal{P}$, with data points $(x_{f1}, x_{m1}), (x_{f2}, x_{m2}), \ldots, (x_{fN}, x_{mN})$, and the child domain as $\mathcal{C}$, with data points $x_{c1}, x_{c2}, \ldots, x_{cN}$. With these notations, one can model the similarity between a child x_c and his/her parents $x_p = (x_f, x_m)$ as

$$s_p(x_p, x_c) = \langle x_p, x_c \rangle_{W_p} \quad (4)$$

where W_p is a $2d \times d$ matrix. This model is called Asymmetric Bilinear Model (ABM) in what follows.

For the ABM model, our verifier is defined as follows:

$$p(y = 1|x_f, x_m, x_c) = \sigma(s_p(x_p, x_c) + b) \quad (5)$$

where σ is the sigmoid function. The parameters $\{W_p, b\}$ are learnt using the following regularized logistic regression objective:

$$\min_{W_p, b} \sum_{i=1}^N \log(1 + exp(-y_i(\langle x_{pi}, x_{ci} \rangle_{W_p} + b))) + \lambda \|W_p\|_* \quad (6)$$

where b is the threshold, and $\|W_p\|_*$ is the trace norm, defined as $\|W_p\|_* = \sum_i \sigma_i$ (the $\sigma_i's$ are the singular values of W_p). With appropriate parameter λ, the trace norm shall force a solution with many singular values of W_p being exactly zero. This allows a more compact representation of the data, thus being useful especially when the original feature space is high-dimensional. Equation 6 is a nonsmooth convex objective and one can use proximal methods to solved it, where at each step the singular values of the standard gradient update are replaced by their soft-threshold versions. See [32] for details on an efficient implementation of this.

Comparing the SBM model and the ABM model, the SBM learns two simple models first (i.e., by learning two $d \times d$ parameter matrices W_f, W_m separately) and then combines them with coefficients β_1 and β_2, while the ABM learns a bigger model at one time (a $2d \times d$ parameter matrix W_p). In other words, the SBM essentially combines two sub-modules (one does the father-child kinship verification and the other for the mother-child relation), which not only makes the learning task easier, but also provides further flexibility to calibrate the outputs of the two sub-modules such that the final prediction (father/mother-child) is as accurate as possible. By contrast, the ABM model tries to do this in one big step, which is much harder especially when the size of dataset is relatively small (less than 2K images for training in our case) for a $2d \times d$ matrix (for $d = 400$, the total number of parameters would be $2 \times 400 \times 400 = 160,000$).

B. Learning A Relative Pairwise Similarity Measure

Note that the SBM model introduced in (1) is a pairwise similarity model without exploiting the dependence structure among parents and child, which can be considered as a limitation. In fact, one can interpret the SBM as a likelihood model, while to better model the similarity between a father and a son for example, one should put it under the context of three subjects - i.e., instead of modeling the marginal pairwise similarity [e.g., p (father is similar to son)], modeling its conditional version [e.g., p (father is similar to son|father, mother, and son)]. One major advantage of this is to allow us to embed various prior knowledge concerning tri-subject groups into the similarity model. In this work, we are particularly interested in the prior knowledge that children may resemble a particular parent more than another [20] - "Jack looks more like his father than his mother" or "John has similar appearance with her mother".

Let us denote the probability that a child looks more like his/her father or his/her mother as p^{fc} and p^{mc} respectively, i.e., '$p^{fc} = 1$' means that a child looks more like his father than his mother. Taking the parents as references, the child is either more like his/her father or more like his/her mother, so we have $p^{fc} + p^{mc} = 1$. We therefore define the two probabilities using the softmax function, based on the pairwise similarity model defined in (1)

$$p^{fc} = \frac{exp(s_f(x_f, x_c))}{exp(s_f(x_f, x_c)) + exp(s_m(x_m, x_c))}$$
$$p^{mc} = \frac{exp(s_m(x_m, x_c))}{exp(s_f(x_f, x_c)) + exp(s_m(x_m, x_c))}. \quad (7)$$

Incorporating these into the SBM model, we obtain the following relative symmetric bilinear model (RSBM)

$$s_f^R(x_f, x_c) = p^{fc} \cdot \langle x_f, x_c \rangle_{W_f}$$
$$s_m^R(x_m, x_c) = p^{mc} \cdot \langle x_m, x_c \rangle_{W_m}. \quad (8)$$

One remaining problem is how to determine these priors. Equation (7) shows that they depend on the parameters W_f and W_m, which suggests a natural iterative procedure - initialize p^{fc} and p^{mc} first, then optimize W_f and W_m in a supervised manner, finally update p^{fc} and p^{mc} again. In this work, we learn W_f and W_m separately using the same trace-norm regularized logistic regression model as that shown in (6).

However, updating p^{fc} and p^{mc} is somewhat subtle - the range of the sigmoid function of (7) is in [0,1], meaning that when one of p^{fc}, p^{mc} reaches 1 the other one must be nearly 0. This is risky, since for the one with 0 probability, the contribution of its corresponding similarity could be cancelled out. To prevent this from happening, we update the new p^{fc}, p^{mc} using a stabilizing term, as follows:

$$p_{new}^{fc} = \alpha p_0^{fc} + (1 - \alpha) p_{cur}^{fc}$$
$$p_{new}^{mc} = \alpha p_0^{mc} + (1 - \alpha) p_{cur}^{mc}, 0 < \alpha < 1 \quad (9)$$

where $\alpha \in (0, 1)$ is a trade-off parameter, and the stabilizing terms p_0^{fc}, p_0^{mc} are initialized to be 0.5 for each sample, and p_{cur}^{fc}, p_{cur}^{mc} are priors calculated according to the W_f or W_m values estimated in the current iteration. In other words, we choose not to trust the currently-estimated similarity prior too much and always regularize it with some fixed stabilizing value. Principally one can optimize the value of α by plugging (9) into the corresponding regularized logistic regression objective function while treating W_f or W_m as a constant, but in our implementation we set it using a cross validation strategy.[2]

The proposed RSBM algorithm is summarized in Algorithm 1.

Algorithm 1 Solving the Relative Symmetric Bilinear Model (RSBM)

Input:
    Training images: $S = \{(x_{fi}, x_{mi}, x_{ci}, y_i)\}_{i=1}^N$;
    Parameters: regularization term λ, iteration number T, and trade-off parameter α
Output:
    Symmetric transformation matrix W_f, W_m;
1:   **Initialization:**
2:     Decompose S into two sets $S_f = \{(x_{fi}, x_{ci}, y_i)\}_{i=1}^N$ and $S_m = \{(x_{mi}, x_{ci}, y_i)\}_{i=1}^N$;
3:     Set $p_0^{fc} = [\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}], p_0^{mc} = [\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}]$;
4:   **For** $L = 1, 2, \ldots, T$ **do**
5:     Estimate W^f and W^m by solving regularized logistic regression objective [c.f., (6)];
6:     Update the pairwise similarity with (8);
7:     Estimate $p_{cur}^{fc}, p_{cur}^{mc}$ using (7);
8:     Update $p_{new}^{fc}, p_{new}^{mc}$ by using (9);
9:     Set $p^{fc} = p_{new}^{fc}; p^{mc} = p_{new}^{mc}$;
10: **end for**
11: **Output** Symmetric transformation W_f and W_m.

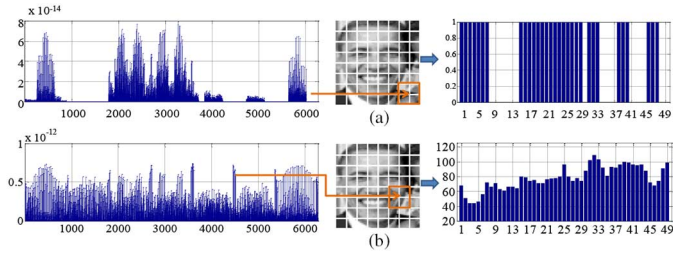[2]In practice, a small value of $\alpha = 0.1$ usually works well.

Fig. 2. Patch selection using (a) the group lasso and (b) the proposed feature selection method. Here, for illustration purpose, the face image in the middle is partitioned into 49 overlapping patches. For (a), the group lasso method directly selects the most discriminative patches by imposing group competition, and the selected groups (patches) are indicated on the right with blue bars, while the corresponding weights of each feature vector in a group is shown in the left histogram; for (b), the discriminative power of each feature (i.e., the weight vector u, see text for details) is first estimated and is shown in the left histogram, while the histogram on the right shows for each patch how many votes it receives, and the first K patches receiving the highest number of votes will be selected.

C. Spatially Voting for Feature Selection

The total number of parameters (i.e., W_p, W_f and W_m) for our kinship verification model grows quadratically with the dimensions of input features, hence performing feature selection is needed. It can be observed that some important genetic characteristics for a kinship relationship are distributed locally in face images, and it is better to learn them by finding the most discriminative local facial regions (patches) with some supervised information. Furthermore, we want to select those most discriminative patches from both parents and the child images simultaneously such that good generalization can be obtained.

One simple way for this is to treat each patch in an image as a group and use existing techniques such as group lasso [33] to select a few groups (patches) such that they give the best prediction accuracy, see Fig. 2(a) or illustration. However, one drawback of this method is that the feature selection is performed at the level of groups (patches), i.e., the features have to be teamed together before competition and this may hurt the flexibility of feature selection. To overcome this, we adopt an alternative strategy - competition before grouping. That is, all features extracted at each location in a given image (c.f., Section III-A on how we extract features) are allowed to freely compete with each other and then select the groups (local regions) in which higher portion of individual features win. Hence our method works in a finer granularity than that of the group lasso. The process of our vote-based feature selection method and group lasso is shown in Fig. 2.

Particularly, our algorithm has two steps. In the first step, we evaluate the discriminative power of each feature of a parent regard to the given child. For this we decompose the triple of (x_f, x_m, x_c) into two pairs of (x_f, x_c) and (x_m, x_c). Then for a pair of father-child features (x_f, x_c), we first concatenate them into a $2d$-dimension vector denoted as a^f, and learn a weight vector u^f with the same dimension using the following sparse l_1 regularized logistic regression objective

$$\min_{u^f} \sum_{i=1}^{N} \log(1 + exp(-y_i \cdot \langle u^f, a_i^f \rangle)) + \gamma \|u^f\|_1 \qquad (10)$$

Fig. 3. Some family image groups of our TSKinFace database, where each group consists of a family triple of a father, a mother, and a child. The first row shows three father-mother-daughter (FM-D) relation families, respectively, and the second row are three father-mother-son (FM-S) relation families, accordingly.

where $y_i = 1$ if the pair is a positive sample and -1 otherwise. Solving this will give us a $2d$-dimension vector u^f with its first half and the second half respectively representing the importance of each feature of the father and the child. The same procedure is repeated for the mother-child pairs and yields a vector u^m.

Now, instead of performing feature selection directly using the information contained in u, we use this to vote the patches of face images and select those patches receiving high votes for face representation. Particularly, after solving the L1 logistic regression, we calculate separately for the parent and child how many votes per patch received from u. Intuitively, the more votes a patch receives, the more important it is.

Fig. 2(b) illustrates this procedure. Since we know the mapping structure between each feature and each patch beforehand, the votes v_k received by the k-th patch can be simply calculated as the sum of weights of u corresponding to this patch, i.e., $v_k = \sum_{j \in k} u_j$, where u_j denotes the j-th element of vector u corresponding to patch k. Note that for a patch of the child image, it would receive votes from the corresponding features of both u^f and u^m, while for a father or a mother patch, its vote comes merely from u^f or u^m accordingly. After voting, we select the first K patches with the highest v_k value for parent and child respectively, where K is set using cross validation over a validation set in our implementation (the best value is usually between 20 and 30 with 49 patches per face.).

As mentioned previously, after patch selection, we collect for each image the selected K patches and encode them with SIFT descriptor, which are further concatenated to form a feature vector x for that face.

IV. THE TSKINFACE DATABASE AND EVALUATION PROTOCOL

To analyze the behavior of the proposed algorithm for tri-subject kinship verification, we constructed a new kinship face database named TSKinFace (Tri-Subject Kinship Face Database). All images in the database are harvested from the internet based on knowledge of public figures family and photo-sharing social network such as flickr.com. During images collecting, we impose no restrictions in terms of pose, lighting, expression, background, race, image quality, etc. Fig. 3 shows some image groups of child-parents pair from our TSKinFace database. This database will be made publicly available online[3] to advance the research and applications related to this topic.

Table I gives a comparison between our TSKinFace database and other existing kinship databases of human faces. It can be seen that our database is characterized by the largest number of

TABLE I
COMPARISON OF OUR TSKINFACE DATABASE AND SOME EXISTING KINSHIP
DATABASES OF HUMAN FACES, WHERE "#GROUPS" REFERS TO THE
NUMBER OF KINSHIP RELATION GROUPS (BLOOD-RELATION FAMILY)
IN THE DATABASE, AND "FAMILY STRUCTURE" REFERS TO THE
EXISTENCE OF FAMILY RELATIONSHIP IN THE DATABASE

Database	#People	#Images	#Groups	Family structure?
CornellKin [12]	300	300	150	NO
UB KinFace [34][14]	400	600	200	NO
KinFaceW-I [15]	1066	1066	533	NO
KinFaceW-II [15]	2000	1000	1000	NO
Family101 [19]	607	14,816	206	YES
Database[18]	–	5400	45	YES
TSKinFace	2589	787	1015	YES

TABLE II
FACE NUMBER INDEX OF EACH FOLD OF THE TSKINFACE DATABASE

Fold	1	2	3	4	5
FM-D	[1,100]	[101,200]	[201,300]	[301,400]	[401,502]
FM-S	[1,102]	[103,204]	[205,306]	[307,408]	[409,513]

people and families. Specifically, the number of families contained in our database is over 20 times more than that of [18] and about 5 times more than that of the Family101 database. These features make our dataset particularly suitable for one-vs-two type kinship verification.

In particular, we are interested in three kinds of child-parents families in real life, i.e., Father-Mother-Daughter (FM-D), Father-Mother-Son (FM-S) and Father-Mother-Son-Daughter (FM-SD). For each type, we collected 274, 285 and 228 family photos respectively, with one photo per family. Using these, we constructed two kinds of family-based kinship relations in the TSKinFace database: Father-Mother-Son(FM-S) and Father-Mother-Daughter(FM-D). The FM-S and the FM-D contain 513 and 502 groups of tri-subject kinship relations (c.f., Fig. 3), respectively. Hence we have 1015 tri-subject groups in our database totally. The families included in our database are diverse in terms of races as well. For FM-S relation, there are 343 and 170 groups of tri-subject kinship relations for Asian and non-Asian, respectively. And for FM-D relation, the numbers for Asian and non-Asian groups are respectively 331 and 171.

Preprocessing: All downloaded images undergo the same geometric normalization prior to analysis: face detected and cropped using our own implemented Viola-Jones detector, rigid scaling and image rotation to place the centers of the two eyes at fixed positions, using the eye coordinates output from an eye localizer [35]; image cropping to 64×64 pixels and conversion to 8 bit gray-scale images. In our experiments, each face image was divided into 7×7 overlapping patches and the size of each patch is 16×16. For each patch, we extracted a 128-dimensional SIFT feature. Except mentioned otherwise, for all experiments described in this work, the SIFT is adopted as our default feature descriptor.

Evaluation Protocol: We design a verification protocol for our database following [13] and [15]: the database is equally divided into five folds such that each fold contains nearly the same number of face groups with kinship relation, which facilitates five-fold cross validation experiments. Table II lists the face number index for the five folds of our TSKinFace database. For face images in each fold, we consider all groups of

face images with kinship relation as positive samples, while the negative samples are a random combination with a child image and two parents images subjected to the constraint that the child was not produced by them. In general, the number of negative samples is much more than that of the positive samples. In our experiments, each couple and child images appeared only once in the negative samples. Hence, the number of positive groups and negative groups are the same.

V. EXPERIMENTS

A. The Tri-Subject Kinship Verification

To the best of our knowledge, there are very few works that tackle the tri-subject kinship verification problem, and it is very difficult to find an existing method directly comparable to ours. We therefore design a naive baseline by concatenating the feature vectors of three visual entities and learning a linear SVM for verification. We denote this method as 'concatenated+SVM'.

Alternatively, one can use any existing state-of-the-art bi-subject kinship verification model to score the similarity between a child and his/her parents separately, and then train a linear SVM over these to make the final prediction [c.f., (3)]. Here two best performers (on the KinFaceW dataset) on bi-subject kinship learning, i.e., neighborhood repulsed metric learning (NRML) [15], and gated autoencoder [25] are adopted as our base models. Furthermore, considering that the similarity modeling is related to metric learning, we also include two classical metric learning algorithms, i.e., Information-theoretic metric learning (ITML) [36] and large margin nearest neighbor classification (LMNN) [37] as the base models.

Although they deal with different problem, the image set based face verification bears some similarities to the problem of tri-subject kinship verification from the respect of methodology, i.e., both involve similarity matching between multiple faces. Hence in this work, we also adopt one of the best performers on the YouTube Face database, i.e., DDML (Discriminative Deep Metric Learning) [38], to score the pairwise similarity between a child and his/her parents, and then train a SVM over these to make the final prediction. Particularly, we train a deep metric learning network with three layers using our own implementation, with the threshold τ, the learning rate μ and regularization parameter λ set to be $3, 10^{-3}, 10^{-2}$, respectively.

Our method is also closely related to Fang *et al.* [19] in that both deal with the family structure. However, since their method is mainly designed for kinship classification, it is not directly comparable to ours. But we follow their ideas to build a linear SVM-based kinship verifier to make the comparison feasible. Particularly, we construct a reconstruction errors-based representation (at the patch level) for each face using sparse group lasso [19], by treating images belonging to the same family as a group.

Finally, we compare several variants of the proposed method in our experiments, as follows.

- With/without feature selection (FS): To investigate the effectiveness of the proposed vote-based feature selection method, for both SBM and ABM models, we evaluate their with/without feature selection versions.
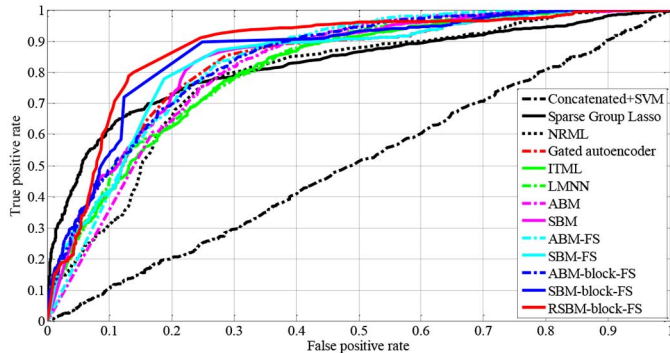
Fig. 4. ROC curves of different methods obtained on the TSKinFace dataset.

Fig. 5. Comparison of FM-S performance of SBM and RSBM, under different number K of patches.

TABLE III
CORRECT VERIFICATION RATES (%) FOR DIFFERENT METHODS ON THE TSKINFACE DATABASE (WHERE "FM-S",'FM-D" DENOTE "FATHER-MOTHER AND SON", "FATHER-MOTHER AND DAUGHTER," RESPECTIVELY)

Method	FM-S	FM-D	avg.
Concatenated+SVM	53.5±0.2381	53.2±0.2037	53.4
Sparse Group Lasso[19]	71.6±0.9644	69.8±0.3485	70.7
NRML [15]	77.0±0.5831	71.4±0.5933	74.2
Gated autoencoder [25]	81.9 ±0.4433	79.6±0.3685	80.8
DDML [38]	82.1 ±1.0357	79.8±0.5879	81.0
ITML [36]	76.6±0.3753	71.4±0.4087	74.0
LMNN [37]	75.4±0.7293	70.3±0.7372	72.9
ABM (*proposed*)	78.5±0.3411	73.2±0.3888	75.9
ABM-FS (*proposed*)	78.6±0.3114	76.9±0.2927	77.8
ABM-block-FS (*proposed*)	83.4±0.2508	81.9±0.3025	82.7
SBM (*proposed*)	82.4±0.3568	78.2±0.4105	80.3
SBM-FS (*proposed*)	82.8±0.2608	79.5±0.2550	81.2
SBM-block-FS (*proposed*)	85.2±0.3031	83.5±0.2985	84.4
RSBM-block-FS(*proposed*)	**86.4±0.4105**	**84.4±0.3601**	**85.4**

- Working at the block level: The proposed method can also be applied at the level of blocks (patches), i.e., selecting the most discriminative patches first, then learning the "covariance" relationship and making verification predictions based on each selected patches, and finally aggregating these meta-decisions through linear SVM for the final verification judgement.

In what follows, a notation like 'RSBM-block-FS' means a Relative Symmetric Bilinear Model (RSBM) with spatially voted feature selection (FS), working at the block level.

Unless otherwise noted, in all experiments we use the following default parameter settings: $\lambda = 5.0$ in (6) (but change to 0.1 if working at the block level); $\gamma = 0.08$ in (10); $\alpha = 0.1$ in (9); and the iteration number T in Algorithm 1 is empirically set to be 5. The influence of some parameters will be investigated in details below, but the exact setting of these parameters is not critical: the method gives similar results over a broad range of settings.

Comparison With the State of the Art Methods: Fig. 4 gives the ROC (Receiver Operating Characteristic) curves of different methods and Table III summarizes the results. One can see from the table that the performance of the baseline SVM algorithm gives an average accuracy of 53.4%, indicating that the one-vs-two type tri-subject kinship verification is a very challenging problem. However, our proposed RSBM model working at the patch level improves this by over 30%, being the best performer among all the compared methods. The closest competitor of our
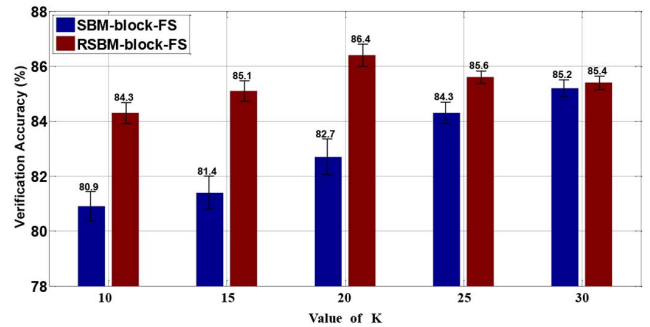
method is the DDML [38], which gives an average accuracy of 81.0% - similar to our method of 'SBM-FS', but with the prior information exploited, our 'RSBM-block-FS' performs better.

Our method also significantly works better than the sparse group lasso based method proposed in Fang *et al.* [19] - one possible explanation is that for a core family group involved only three subjects, the assumption made in [19] that an image of a child should be best reconstructed by face images in his/her own family is too strong, although it is reasonable under their situation where dozens of face images per family are available.

Thirdly, we see that simply adopting state of the art metric learning methods for tri-subject kinship verification is not a good choice. This is partly due to the fact that these methods fail to model the dependence structure among the three visual entities. By contrast, the proposed RSBM model effectively exploits such priors during several stages of the verification pipeline (e.g.,similarity modeling, feature selection) and achieves the best verification performance.

Last but not least, it is interesting to note that the gender plays a significant role in kinship verification - in all the cases tested, the verification rates on "FM-S" are consistently higher than those on "FM-D". One possible reason is that the appearance variations appeared in a female subject (daughter) are more complex than those in a male subject (son). This seems to be in accordance with earlier psychological research results [39] that the kin recognition signal is less evident from daughters than from sons.

The Importance of Prior Knowledge: Fig. 5 compares in detail the FM-S performance of the SBM model with/without exploiting the prior knowledge about the relative difference of a child to his/her parents, as a function of the number of patches selected for each face. One can see that when the number of patches selected for verification is relatively small, the RSBM method significantly performs better than the SBM method. For example, using only 20 patches, the accuracy of RSBM reaches an verification accuracy of 86.4%, 3.7% higher than that of the SBM model. This highlights the benefits of exploiting prior knowledge for complex kinship verification. To further illustrate this, we visualize the prior knowledge learned by multiplying it elementwise by the image: see Fig. 6 for an example. We can observe that some children do look more like his/her father than mother, or vice versa, and such information is effectively captured and utilized by our model.
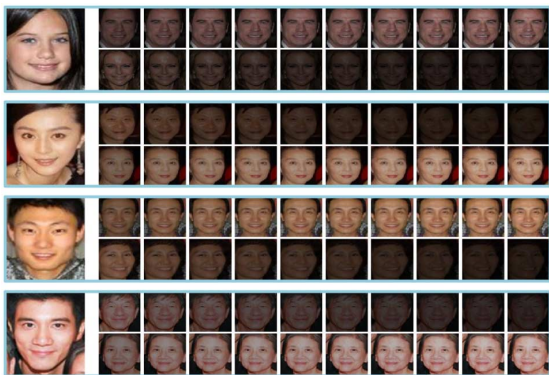
Fig. 6. Illustration of the learned prior knowledge for four families. In each family, the image on the left is the input child image, and the two rows of images on the right are the parents images multiplied by the respective learnt prior density in 10 iterations (progressively from left to right)—the higher the probability the lighter the pixel value.
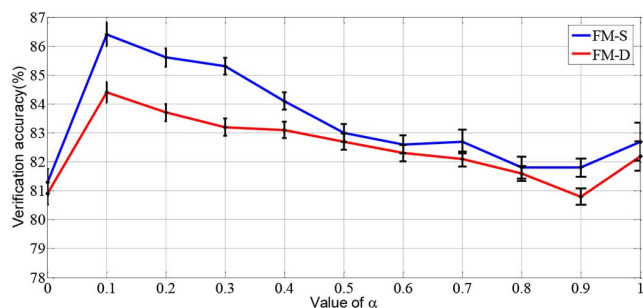

Fig. 7. Average performance of "RSBM-block-FS" as a function of the amount of stabilization α.
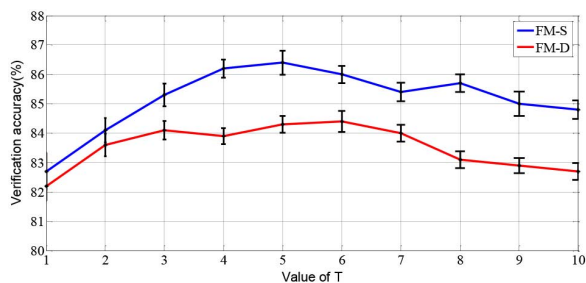

Fig. 8. Average verification accuracy of "RSBM-block-FS" as a function of the number of iteration T.

Fig. 7 shows the average verification accuracy of our RSBM model as a function of the stabilizing term α [c.f., (9)]. We can see that the RSBM model obtains the best performance when $\alpha = 0.1$ for both FM-S and FM-D. In general, good performance could be obtained by setting the value of α between 0.05 and 0.3.

Fig. 8 shows the performance curve of the RSBM model as a function of the number of iterations. We can see that the performance of the RSBM model boosts to its highest value only after a few iterations. In practice we would recommend to set $T = 5$ to avoid overfitting.

Effectiveness of Spatially Voting for Feature Selection: To verify this, we compare our feature selection scheme with group lasso (GL) - both aim to automatically figure out a set of patches
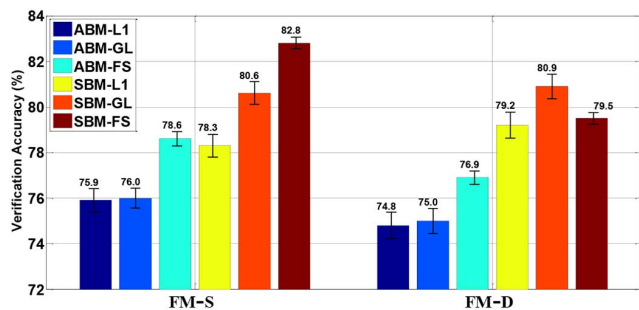

Fig. 9. Correct verification rates (%) for different feature selection methods on the TSKinFace database (,"FS" denotes our vote based feature selection method, "L1" denotes lasso, and "GL" denotes group lasso).
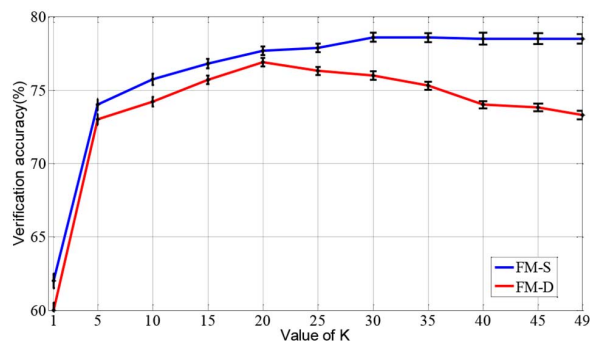

Fig. 10. Influence of the number of patches K selected on the verification rates.

from face images for kinship verification. Particularly, we formulate the problem as the sparse group lasso penalized logistic regression in which the groups are defined as the patches. For a fair comparison, we set the parameter that controls the group weights as 0.88 for FM-D and 0.85 for FM-S, obtaining the same number of selected patches as that in our vote-based feature selection method. As the baseline we select the l_1 norm lasso algorithm, which performs the feature selection without using any spatial information.

Fig. 9 gives the results. One can see that the proposed spatially voted feature selection scheme ("FS") performs better than the group lasso "GL", on average improving the performance by about 2.3% and 0.4%, respectively on both tasks, while the simple lasso method performs the worst.

To answer the question of how many patches it needs to be selected, we investigate the effect of the parameter K (the number of patches selected) on the performance. Fig. 10 shows verification rate as a function of the number of patches selected for each face, with the ABM model as the verifier. One can see that the performance boots from about 60.0% to over 73.0% with only 5 patches. The performance increases with more patches added until 20 patches are selected, and the improvement is not evident after that for the FM-S verification. While for the FM-D verification, the number of selected patches is better to be kept less than 20 so as to reduce the possible influence of noise.

Influence of Randomness in Negative Sample Generations: In the previous experiments the negative samples are randomly generated by combining child and parents from different families but there is only one for each. We now investigate the impact of such randomness. Particularly, we first randomly generate

Fig. 11. Illustration of some samples in the easy negative set (the top two rows) and the hard negative set (the bottom two rows).

TABLE IV

CORRECT VERIFICATION RATES (%) FOR DIFFERENT METHODS ON THE 5TH FOLD OF TSKINFACE DATABASE WITH 10 DIFFERENT NEGATIVE SAMPLES SETS, WHERE "EASY" REPRESENTS THOSE SAMPLES WHICH ARE CLASSIFIED BY "LMNN" CORRECTLY AND "HARD" DENOTES THOSE SAMPLES WHICH ARE CLASSIFIED BY "LMNN" INCORRECTLY

Method	Easy	Hard	avg.
LMNN [37]	100.0	0.0	50
Concatenated+SVM	46.3±0.0224	38.1±0.0173	42.2±0.0176
NRML [15]	78.4±0.0046	51.4±0.0256	64.9±0.0053
DDML [38]	77.6±0.0083	55.1±0.0163	66.4±0.0086
ABM-block-FS (*proposed*)	87.3±0.0041	54.0±0.0146	70.7±0.0030
SBM-block-FS (*proposed*)	92.5±0.0037	57.3±0.0143	74.9±0.0036
RSBM-block-FS (*proposed*)	**94.7±0.0040**	**63.1±0.0142**	**78.9±0.0032**

a large negative samples pool containing both the FM-D and the FM-S negative relationship. Although there are many ways to distinguish less distinct negative samples (i.e., hard samples with features of child and parents not very different) from those very distinct negative samples (i.e, easy ones with features of child and parents quite different), for example, by simply measuring the similarity between the child and parents, we choose the "LMNN" method [37] here. That is, those samples correctly classified by the "LMNN" method as negative are categorized as "easy" ones, otherwise as "hard". We use this criterion to randomly select 2,000 samples from the negative pool, with 1,000 each for the "easy" and the "hard" category respectively. Some of these samples are illustrated in Fig. 11. For experiment we equally split those samples into 10 sets, and run the model trained on the fifth fold of our TSKinFace database over them.

Table IV gives the results. One can see that all the methods investigated here work consistently and significantly better on the easy set than on the hard set, indicating that it makes sense to distinguish the two sets based on the outputs of the "LMNN" method, while our "RSBM-block-FS" method demonstrates the highest robustness against this random confusion. The table also reveals that although some of the methods work well on the easy negative samples, the performance on the hard ones are not satisfactory in general (with accuracy less than 70.0%), showing that further research is needed on this topic.

Computational Complexity: We now briefly analyze the computational complexity of the RSBM method, which involves T iterations. In each iteration we solve a regularized logistic regression problem and make the estimation of the weights of p^{fc} and p^{mc}. To solve the regularized logistic
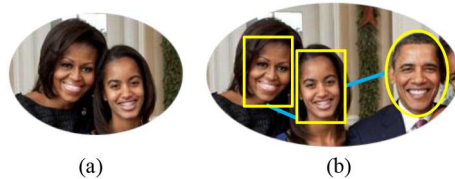

Fig. 12. When the images of a second parent is available, the traditional one-versus-one type bi-subject kin verification problem can be reformulated as a one-versus-two one. In this example, once the father/mother-daughter (FM-D) relationship is established for the three subjects shown on the right, one can safely infer that the mother-daughter (MD) kinship is validated for the subjects shown on the left.

regression problem, we use a fast implementation [32] with its computational complexity $O(d^3 N/g^2)$, where g is the iteration counter, while the computational complexity of the estimating part is $O(N)$. Hence the total computational complexity of our proposed RSBM is $O(d^3 N/g^2 T) + O(NT)$.

B. Enhancing Bi-Subject Kinship Verification

Intuitively, having more information about one's parents is potentially useful to improve the performance of bi-subject kinship verification. In order to verify this hypothesis, another series of experiments are conducted. This is similar to the traditional bi-subject verification in that four types of kinship relations will be evaluated, i.e., Father-Son(FS), Father-Daughter (FD), Mother-Son (MS), and Mother-Daughter (MD). However, the key difference lies in that we are now given a triple including two parents and a child as a test sample. In other words, we are interested in, for example, whether the information about one's father is useful to verify the Mother-Daughter (MD) relation.

One simple way for this is to reformulate the bi-subject kinship verification problem as a tri-subject problem, since once a FM-D relationship is established, a FD and a MD relationship *must* be established as well, see Fig. 12 for an example. For a bi-subject verification problem shown in Fig. 12(a), one can treat it as a tri-subject problem shown in Fig. 12(b) when the father's information is available, and use that result to answer the question of one-vs-one kinship verification.

Table V compares the results of these two approaches for bi-subject kinship verification. One can see that exploiting more information about one's parents is indeed beneficial. Particularly, the performance of the mother-son (MS) verification is improved significantly from 72.2% to 77.0% using the SVM-based NRML baseline method, while that of the father-daughter (FD) verification is improved from 80.5% to 84.4% using our RSBM model, and t-test analysis shows that this improvement is statistically significant. Particularly, the stars in Table V indicate whether the improvement of performance for triple inputs is significant compared to that when only two subjects are available. For example, the accuracy of 'SBM-block-FS' for father-son verification (FS) is 83.0%, but if the information about the mother is known, this improves to 85.2% (as shown in the column of 'FM-S'), which is significantly better than that of FS according to the t-Test and hence a star is marked, otherwise there is no star.

It is well known that a problem like MS or FD verification is quite difficult due to the different genders of two subjects to be verified. Our method essentially provides a new solution to this,

TABLE V
CORRECT RATES (%) OF DIFFERENT METHODS FOR BI-SUBJECT KINSHIP VERIFICATION WITH TRIPLE INPUTS (COLUMN 2 AND 5) AND PAIR INPUTS (COLUMN 3, 4, 6, AND 7, "*" DENOTES THAT THE RESULT (P-VALUES) OF $t-$ TEST FOR THE PERFORMANCE COMPARISON BETWEEN PAIR INPUTS AND TRIPLE INPUTS VERIFICATION IS LESS THAN 0.05).

Method	FM-S	FS	MS	FM-D	FD	MD
Sparse Group Lasso[19]	71.6±0.9644	69.1±0.6093	68.7±1.2204	69.8±0.3485	66.8±0.4627(*)	67.9±0.5977
NRML [15]	77.0±0.5831	74.8±0.7279(*)	72.2±0.3360(*)	71.4±0.5933	70.0±0.6716(*)	71.3±0.5853
Gated autoencoder [25]	81.9±0.4433	79.9±0.6790(*)	78.5±0.5963(*)	79.6±0.3686	74.2±0.3170(*)	76.3±0.2296(*)
ITML [36]	76.6±0.3753	75.6±0.3866(*)	72.1±0.3330(*)	71.4±0.4087	70.5±0.4000(*)	70.7±0.4435(*)
LMNN [37]	75.4±0.7293	72.7±0.7305	71.5±0.7455(*)	70.3±0.7372	69.8±0.7243(*)	70.1±0.3846
ABM-block-FS (*proposed*)	83.4±0.2508	83.0±0.5558	82.8±0.5037	81.9±0.3025	80.5±0.4301	81.1±0.4003
SBM-block-FS (*proposed*)	85.2±0.3031	83.0±0.5558(*)	82.8±0.5037(*)	83.5±0.2985	80.5±0.4301(*)	81.1±0.4003(*)
RSBM-block-FS(*proposed*)	**86.4±0.4105**	83.0±0.5558(*)	82.8±0.5037(*)	**84.4±0.3601**	80.5±0.4301(*)	81.1±0.4003(*)

TABLE VI
CORRECT RATES (%) OF HUMAN BEINGS AND OUR METHOD ON THE FM-S SUBSET OF THE TSKINFACE DATABASE

Method	FM-S	FS	MS
A	N/A	77.3±2.1927	74.2±1.6592
B	79.9±1.6362	N/A	N/A
SBM(*proposed*)	81.6±0.2875	N/A	N/A
SBM-FS(*proposed*)	81.9±0.3479	N/A	N/A
SBM-block-FS(*proposed*)	82.4±0.6419	N/A	N/A
RSBM-block-FS(*proposed*)	85.4±0.1789	N/A	N/A

TABLE VII
CORRECT RATES (%) OF HUMAN BEINGS AND OUR METHOD ON THE FM-D SUBSET OF THE TSKINFACE DATABASE

Method	FM-D	FD	MD
A	N/A	73.5±1.2042	75.5±1.2942
B	79.2±1.4415	N/A	N/A
SBM(*proposed*)	79.2±0.4131	N/A	N/A
SBM-FS(*proposed*)	80.0±0.6127	N/A	N/A
SBM-block-FS(*proposed*)	81.4±1.0354	N/A	N/A
RSBM-block-FS(*proposed*)	83.0±0.8000	N/A	N/A

Fig. 13. Illustration of the construction of a new test dataset with different lighting conditions using the face images from the Family101 database. For each family, there are several face images per subject (upper row: parents; lower row: two children). We randomly select one image from two parents and one child to construct a triple-item group, such that all the images from the same group do not come from the same photo.
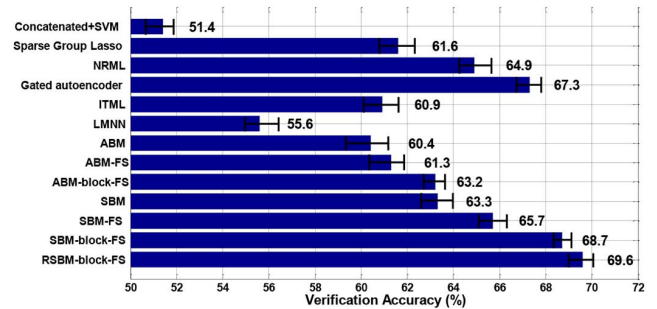

Fig. 14. Correct verification rates (%) for different methods on the Family101 subset database.

and we consider it as one of the major motivations to study the tri-subject kinship verification problem.

C. Comparisons With Human Beings in Kinship Verification

To investigate human beings' performance on the kinship verification problem, we randomly selected 100 groups of cropped grayscale face samples, including 50 positive groups and 50 negative ones. Then we presented these to 10 human observers with ages of 20 to 40 years old to ask their opinions about the kinship relation. These human observers did not receive any training on how to verify kinship from facial images before the experiment, and will completely rely on their own knowledge accumulated to answer the questions.

Particularly, we conduct two parts of tests on kinship verification. For the first part, 100 child-parent pairs (one-vs-one) are shown to human observers ("A"), and for the second part, 100 child-parents groups (one-vs-two) are presented to these observers ("B"). Obviously, these two types of testing are respectively corresponding to the problem of bi-subject and tri-subject kinship verification. We repeated this procedure two times, one for the FM-S subset and other for the FM-D subset, both from our TSKinFace database. We also run the same experiments using our SBM method for comparison.

Tables VI and VII give the results. One can see that "B" can obtain better performance than "A" on the two subsets, which indicates that human beings are able to combine the information from both parents to make better kinship judgement. For example, the performance of the mother-son (MS) verification is 74.2%, but if the face image of one's father is also available,

the performance increases to 79.9%. Moreover, it is worth mentioning that our proposed SBM methods achieve higher verification accuracies than "B".

D. Robustness Under Different Lighting Conditions

Since all the face images in a family in our database are extracted from the same photo, it could introduce unnecessary bias in learning. In order to investigate the behavior of our algorithm when encounters face images from completely different lighting conditions, we construct a new dataset based on the Family101 [19].

Particularly, we manually selected 48 families from 206 nuclear families of Family101, with the following conditions: 1) each family contains four members, i.e., father, mother and two children, and 2) at least 3 face images exist for each family member. We then cropped these images to 64×64 pixels and converted them to 8 bit gray-scale, divided them into 7×7 overlapping patches and extracted SIFT features. Fig. 13 gives some examples of the preprocessed images.

TABLE VIII
CORRECT VERIFICATION RATES (%) FOR DIFFERENT METHODS ON THE TSKINFACE DATABASE (WHERE "FS-M", "MS-F", "FD-M", AND "MD-F" DENOTE "FATHER-SON AND MOTHER", "MOTHER-SON AND FATHER", "FATHER-DAUGHTER AND MOTHER", AND "MOTHER-DAUGHTER AND FATHER", RESPECTIVELY)

Method	FS-M	MS-F	FD-M	MD-F	avg.
Concatenated+SVM	53.5±0.2066	53.8±0.1337	53.0±0.1732	53.3±0.2523	53.4
Sparse Group Lasso[19]	69.9±0.4927	70.3±0.7489	68.6±0.5350	68.9±0.5070	69.4
NRML [15]	75.6±0.5931	75.8±0.3788	70.8±0.9266	70.2±0.3189	73.1
Gated autoencoder [25]	79.7±0.4077	80.9±0.4023	78.4±0.4022	79.2±0.4133	79.6
ITML [36]	73.8±0.4921	74.6±0.2564	70.1±0.3727	70.0±0.4950	72.1
LMNN [37]	71.9±0.2372	72.6±0.7213	70.5±0.5129	69.5±0.4232	71.1
ABM (*proposed*)	78.0±0.4379	78.7±0.3117	73.1±0.4190	73.5±0.5103	75.8
ABM-FS (*proposed*)	77.9±0.3104	78.0±0.4235	75.1±0.3008	76.5±0.4730	76.9
ABM-block-FS (*proposed*)	81.3±0.2791	81.8±0.3984	80.4±0.3604	80.7±0.3309	81.1
SBM (*proposed*)	79.2±0.4113	80.4±0.4103	76.8±0.4216	77.1±0.4010	78.4
SBM-FS (*proposed*)	81.0±0.3716	81.7±0.3002	78.7±0.3919	79.0±0.2637	80.1
SBM-block-FS (*proposed*)	**82.9±0.1841**	**83.9±0.4197**	**81.9±0.1071**	**81.8±0.2157**	82.6

Then, to construct tri-subject groups for our tri-subject kinship verification, we do the following iterations for each of the selected family:

1) select one image among 3 images from the father's;
2) select one image among 3 images from the mother's; and
3) select one image among 6 images from the two children's.

This will give us $3 \times 3 \times 6 = 54$ different groups per family. In the experiments, we follows the four-fold cross validation protocol, which means that for each round of evaluation, 36 families will be used for training while the remaining 12 for testing. In other words, in training we will have 36 families $\times 54$groups/family $= 1944$ groups in total. And no any three face images in each group appear in one photos before. This will suppress the bias for the positive samples to have similar lighting conditions as much as possible.

Fig. 14 gives the results. One can see that all the methods are influenced by the illumination changes introduced in the dataset. However, the proposed 'RSBM-block-FS' performs the best among the compared ones, about 18.2% higher than the baseline algorithm in terms of accuracy. The table also reveals that by replacing the pairwise bilinear similarity with the proposed relative similarity measure, one can improve the performance from 68.7% ('SBM-block-FS') to 69.6% ('RSBM-block-FS').

E. Other Forms of Tri-Subject Kinship Verification

In previous sections we focus on the child-parents type tri-subject kinship verificaiton, but the same method could also be applied to verify other types of tri-subject kin relations, i.e., Father/Son-Mother (FS-M), Mother/Son-Father (MS-F), Father/Daughter-Mother (FD-M), Mother/Daughter-Father (MD-F). For example, the task of FS-M is to verify whether a valid kin relation could be established between a mother and a father-son pair, given their face images.

For this series of experiments, we adopt the same 5 cross-validation evaluation protocol introduced in Section 4 for each type of verification, with the only exception that images in each fold are partitioned according to the type of kinship of interest. We use SIFT features for face representation and follows the same parameter settings as previous experiments.

Table VIII gives results. It can be seen that the performance obtained here for different methods generally decreases by about 2-3% compared to that in the child-parents verification (c.f., Table III). One possible explanation could be this: for a mixed one-vs-two relation, taking the FS-M relation for

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

Label	FS	FD	MS	MD	avg.
Polito [41]	85.30	**85.80**	**87.50**	86.70	86.30
LIRIS [41]	83.04	80.63	82.30	84.98	82.74
ULPGC [41]	71.25	70.85	58.52	80.89	70.01
BIU [41]	86.90	76.48	73.89	79.75	79.25
NUAA(proposed)[41]	86.25	80.64	81.03	83.93	82.96
M-NUAA(proposed)	**87.84**	85.47	86.16	**87.50**	**86.74**
SILD(LBP)[41]	78.22	69.40	66.81	70.10	71.13
SILD(HOG)[41]	80.46	72.39	69.82	77.10	74.94

example, one has to decompose the triples of (x_f, x_c, x_m) into two pairs of (x_f, x_m) and (x_c, x_m), and learn the pairwise similarity respectively. But the appearance similarity between a father and a mother is more difficult to learn, compared to that between a child and a mother. However, even under such a scenario, it can be seen that our proposed method (SBM-block-FS) obtains the best verification performance.

One interesting question naturally arises here is whether the appearance between a father and a mother is really similar to each other? Possibly not, because a father and a mother have different gender and have no blood relationship. But a positive father-mother pair is actually spouses who have lived together under the same living environment for a period of time, which, according to some research [40], could make their appearance look more similar to each other than to others. While the size of our database is still not big enough to support this, it deserves more attention in our future research.

F. The Bi-Subject Kinship Verification

In the final series of experiments, we briefly evaluate the performance of the proposed method on the task of the bi-subject kinship verification. Particularly, we do this on two largest datasets for bi-subject verification: KinFaceW-I [15] and KinFaceW-II [15]. The KinFaceW-I database consists of 156 FS (Father-Son), 134 FD (Father-Daughter), 116 MS (Mother-Son) and 127 MD (Mother-Daughter) pairs, while the KinFaceW-II contains 250 pairs of these bi-subject kin relations each. The major difference between KinFaceW-I and KinFaceW-II lies in that each pair of faces in KinFaceW-I comes from the same photo while from different photos in KinFaceW-II. Hence the latter one is easier than the former.

We follow the evaluation protocol as proposed in [41]. Table IX and Table X give the baseline and other latest

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

Label	FS	FD	MS	MD	avg.
Polito [41]	84.00	82.20	84.80	81.20	83.10
LIRIS [41]	**89.40**	83.60	**86.20**	85.00	86.05
ULPGC [41]	85.40	75.80	75.60	81.60	80.00
BIU [41]	87.51	80.82	79.78	75.63	80.94
NUAA(proposed) [41]	84.40	81.60	82.80	81.60	82.50
M-NUAA(proposed)	88.40	**86.20**	86.00	**85.20**	**86.45**
SILD(LBP)[41]	78.20	70.00	71.20	67.80	71.80
SILD(HOG) [41]	79.60	71.60	73.20	69.60	73.50

state-of-the-art results, where the performance of the methods are directly cited from the corresponding paper. It can be seen that our original method (i.e., "NUAA") obtains rank 3 with a simple bilinear model and without using any other features except SIFT (while both the top two methods combine several kinds of features).

It can be conjectured that combining multiple feature information could be beneficial to the performance. Hence in the next round of experiments we add two other features (i.e., the C-SVDD features [42] and TPLBP [43]) and fuse them at the decision level. This multiple feature version is denoted as "M-NUAA" in Table IX and Table X. One can see that the improved "M-NUAA" method achieves results better than or comparable to the state of the art methods on both bi-kinship datasets, in terms of average performance (last column in both tables). Note that our algorithm is the first one designed for handle the one-vs-two tri-kinship problem and others do not, which is actually the main advantage of the proposed method: it can be thought of as a framework which can encompass any algorithm of bi-subject kinship verification for tri-subject kinship verification, while effectively incorporating useful prior knowledge.

VI. CONCLUSION

In this work, we made the first attempt to investigate the tri-subject kinship verification problem extensively. Instead of using information from a single parent, we exploit information from both parents to learn the kinship relationship between them and their child, which is arguably one of the most important relationships formed in a family. For this we proposed a novel relative symmetric bilinear model (RSBM) and a spatially voted feature selection method, both incorporate prior knowledge about the dependence structure between a child and his/her two parents. Furthermore, we collected a new kinship face database characterized by over 1,000 groups of triples, on which we show that our method achieves state of the art verification accuracy. Our experimental results also reveal that the proposed method could be used to significantly boost the performance of bi-subject kinship verification when the information about both parents is available. Additionally, we show that our method can be applied with encouraging performance on other types of tri-subject kinship verification such as Father/Son-Mother verification, and on the traditional one-vs-one kinship problem.

Future works include further improvement based on exploiting other types of prior knowledge and learning multiple complementary features to better represent the discriminative

information that is useful for our task. We also plan to extend our framework to handle more general family structure.

REFERENCES

[1] Y.-G. Jiang, J. Wang, X. Xue, and S.-F. Chang, "Query-adaptive image search with hash codes," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 442–453, Feb. 2013.

[2] X. Benavent, A. Garcia-Serrano, R. Granados, J. Benavent, and E. de Ves, "Multimedia information retrieval based on late semantic fusion approaches: Experiments on a Wikipedia image collection," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2009–2021, Dec. 2013.

[3] E. Spyromitros-Xioufis, S. Papadopoulos, I. Kompatsiaris, G. Tsoumakas, and I. Vlahavas, "A comprehensive study over VLAD and product quantization in large-scale image retrieval," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1713–1728, Oct. 2014.

[4] Z. Ma, Y. Yang, N. Sebe, K. Zheng, and A. G. Hauptmann, "Multimedia event detection using a classifier-specific intermediate representation," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1628–1637, Nov. 2013.

[5] O. Zoidi, A. Tefas, N. Nikolaidis, and I. Pitas, "Person identity label propagation in stereo videos," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1358–1368, Aug. 2014.

[6] G. Wang, A. Gallagher, J. Luo, and D. Forsyth, "Seeing people in social context: Recognizing people and social relationships," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 169–182.

[7] M. Shao, S. Xia, and Y. Fu, "Identity and kinship relations in group pictures," in *Human-Centered Social Media Analytics*. New York, NY, USA: Springer, 2014, pp. 175–190.

[8] X. Zhou, J. Lu, J. Hu, and Y. Shang, "Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 725–728.

[9] Z. Xu, Y. Zhang, and L. Cao, "Social image analysis from a non-IID perspective," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1986–1998, Nov. 2014.

[10] G. Guo and X. Wang, "Kinship measurement on salient facial features," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2322–2325, Aug. 2012.

[11] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," *CoRR*, vol. abs/1406.4773, 2014 [Online]. Available: http://arxiv.org/abs/1406.4773

[12] R. Fang, K. D. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," in *2010 IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1577–1580.

[13] S. Xia, M. Shao, and Y. Fu, "Kinship verification through transfer learning," in *Proc. 22nd Int. Joint Conf. Artificial Intell.*, 2011, vol. 3, pp. 2539–2544.

[14] S. Xia, M. Shao, and Y. Fu, "Toward kinship verification using visual attributes," in *Proc. IEEE Int. Conf. Pattern Recog.*, Nov. 2012, pp. 549–552.

[15] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 331–345, Feb. 2014.

[16] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Like father, like son: Facial expression dynamics for kinship verification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1497–1504.

[17] H. Yan, J. Lu, W. Deng, and X. Zhou, "Discriminative multi-metric learning for kinship verification," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1169–1178, Jul. 2014.

[18] M. Ghahramani, W.-Y. Yau, and E. K. Teoh, "Family verification based on similarity of individual family members facial segments," *Mach. Vis. Appl.*, vol. 25, no. 4, pp. 919–930, 2014.

[19] R. Fang, A. C. Gallagher, T. Chen, and A. Loui, "Kinship classification by modeling facial feature heredity," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 2983–2987.

[20] A. Alvergne, C. Faurie, and M. Raymond, "Differential facial resemblance of young children to their parents: Who do children look like more?," *Evolution Human Behavior*, vol. 28, no. 2, pp. 135–144, 2007.

[21] Q. Xiaoqian, T. Xiaoyang, and C. Songcan, "Tri-subjects kinship verification: Understanding the core of a family," in *Proc. Int. Conf. Mach. Vis. Appl.*, to be published.

[22] M. F. Dal Martello and L. T. Maloney, "Where are kin recognition signals in the human face?," *J. Vis.*, vol. 6, no. 12, p. 2, 2006.

[23] L. M. DeBruine *et al.*, "Kin recognition signals in adult faces," *Vis. Res.*, vol. 49, no. 1, pp. 38–43, 2009.

[24] X. Zhou, J. Hu, J. Lu, Y. Shang, and Y. Guan, "Kinship verification from facial images under uncontrolled conditions," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 953–956.

[25] A. Dehghan, E. G. Ortiz, R. Villegas, and M. Shah, "Who do I look like? determining parent-offspring resemblance via gated autoencoders," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 1757–1764.

[26] N. Kohli, R. Singh, and M. Vatsa, "Self-similarity representation of Weber faces for kinship classification," in *Proc. IEEE 5th Int. Conf. Biometrics: Theory, Appl. Syst.*, Sep. 2012, pp. 245–250.

[27] H. Yan, J. Lu, and X. Zhou, "Prototype-based discriminative feature learning for kinship verification," *IEEE Trans. Cybern.*, to be published.

[28] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan, "Large margin multimetric learning for face and kinship verification in the wild," in *Proc. ACCV*, 2014, pp. 252–267.

[29] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *Int. J. Comput. Vis.*, vol. 109, no. 1–2, pp. 74–93, 2014.

[30] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding kin relationships in a photo," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1046–1056, Aug. 2012.

[31] N. Syed and B. K. P. P. Q. Shareq *et al.*, "Understanding familial relationship in an image," *Int. J. Sci. Res. Educ.*, vol. 2, no. 06, pp. 1037–1045, 2014.

[32] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proc. 26th Annual Int. Conf. Mach. Learn.*, 2009, pp. 457–464.

[33] L. Meier, S. Van De Geer, and P. Bühlmann, "The group lasso for logistic regression," *J. Royal Statist. Soc.: Ser. B (Statist. Methodology)*, vol. 70, no. 1, pp. 53–71, 2008.

[34] M. Shao, S. Xia, and Y. Fu, "Genealogical face recognition based on UB KinFace database," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2011, pp. 60–65.

[35] X. Tan, F. Song, Z.-H. Zhou, and S. Chen, "Enhanced pictorial structures for precise eye localization under uncontrolled conditions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 1621–1628.

[36] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 209–216.

[37] K. Weinberger, J. Blitzer, and L. Saul, "Distance metric learning for large margin nearest neighbor classification," *Adv. Neural Inf. Process. Syst.*, vol. 18, pp. 18–1473, 2006.

[38] J. Hu, J. Lu, and Y.-P. Tan, "Discriminative deep metric learning for face verification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 1875–1882.

[39] S. M. Platek *et al.*, "Reactions to children's faces: Males are more affected by resemblance than females are, and so are their brains," *Evolution Human Behavior*, vol. 25, no. 6, pp. 394–405, 2004.

[40] B. W. Domingue, J. Fletcher, D. Conley, and J. D. Boardman, "Genetic and educational assortative mating among us adults," in *Proc. Nat. Academy Sci.*, 2014, vol. 111, no. 22, pp. 7996–8000.

[41] J. Lu *et al.*, "The FG 2015 kinship verification in the wild evaluation," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recog.*, May 2015, pp. 1–7.

[42] D. Wang and X. Tan, "Centering SVDD for unsupervised feature representation in object classification," in *Neural Information Processing*. New York, NY, USA: Springer, 2013, pp. 376–383.

[43] L. Wolf *et al.*, "Descriptor based methods in the wild," in *Proc. Workshop Faces "Real-Life" Images: Detection, Alignment, Recog.*, 2008, pp. 4–12.
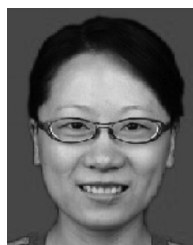
Xiaoqian Qin is currently working toward the Ph.D. degree at the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

She is a Lecturer with Huaiyin Normal University, Huaiyin, China. Her research interests include computer vision and machine learning.
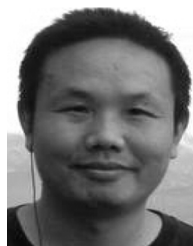
Xiaoyang Tan received the Ph.D. degree from Nanjing University, Nanjing, China, in 2005.

He is currently a Professor with the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China. From 2006 to 2007, he was a Postdoctoral Researcher with the LEAR Team, INRIAR Rhone-Alpes, Grenoble, France. He has authored or coauthored over 40 conference and journal papers. His research interests include face recognition, computer vision, machine learning, and artificial intelligence.

Songcan Chen received the B.S. degree in mathematics from Hangzhou University, Zhejiang, China, in 1983, the M.S. degree in computer applications from Shanghai Jiaotong University, Shanghai, China, in 1985, and the Ph.D. degree in communication and information systems from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 1997.

He joined the NUAA in 1986. Since 1998, he has been a Full-Time Professor with the Department of Computer Science and Technology, NUAA. He has authored or coauthored over 200 scientific peer-reviewed papers. His current research interests include pattern recognition, machine learning, and neural computing.