Pattern Recognition 75 (2018) 15–24

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Video-based kinship verification using distance metric learning

Haibin Yan[a,*], Junlin Hu[b]

[a] School of Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798, Singapore

ARTICLE INFO

Article history:
Received 1 September 2016
Revised 7 January 2017
Accepted 1 March 2017
Available online 7 March 2017

Keywords:
Kinship verification
Metric learning
Face recognition
Video-based

ABSTRACT

In this paper, we investigate the problem of video-based kinship verification via human face analysis. While several attempts have been made on facial kinship verification from still images, to our knowledge, the problem of video-based kinship verification has not been formally addressed in the literature. In this paper, we make the two contributions to video-based kinship verification. On one hand, we present a new video face dataset called Kinship Face Videos in the Wild (KFVW) which were captured in wild conditions for the video-based kinship verification study, as well as the standard benchmark. On the other hand, we employ our benchmark to evaluate and compare the performance of several state-of-the-art metric learning based kinship verification methods. Experimental results are presented to demonstrate the efficacy of our proposed dataset and the effectiveness of existing metric learning methods for video-based kinship verification. Lastly, we also evaluate human ability on kinship verification from facial videos and experimental results show that metric learning based computational methods are not as good as that of human observers.

1. Introduction

Kinship verification from human faces in a relatively new problem in biometrics in recent years. The key motivation of this research topic is from the research observations in results in psychology and cognitive sciences [1–4] where human faces convey an important cue for kin similarity measure because children usually look like their parents. Verifying human kinship relationship has several potential applications such as image annotation, family album organization, social media mining, and missing children searching. Over the past few years, a number of kinship verification methods have been proposed in the literature, which aims to present effective computational models to verify human kinship relations via facial image analysis [5–17]. While these methods have achieved some encouraging performance [5–13,15,18], it is still challenging to develop discriminative and robust kinship verification approached for real-world applications, especially when face images are captured in unconstrained environments where large variations of pose, illumination, expression, and background occurs.

Most existing kinship verification methods determine human kinship relationship from still face images. Due to the large variations of human faces, a single still image may not be discriminative enough to verify human kin relationship. Compared to a single image, a face video provides more information to describe the appearance of human face. It can capture the face of the person of interest from different poses, expressions, and illuminations. Moreover, face videos can be much easier captured in real applications because there are extensive surveillance cameras installed in public areas. Hence, it is desirable to employ face videos to determine the kin relations of persons. However, it is also challenging to exploit discriminative information of face videos because intra-class variations are usually larger within a face video than a single sill image.

In this paper, we investigate the problem of video-based kinship verification via human face analysis. Specifically, we make the two contributions to video-based kinship verification. On one hand, we present a new video face dataset called Kinship Face Videos in the Wild (KFVW) which were captured in wild conditions for the video-based kinship verification study, as well as the standard benchmark. On the other hand, we employ our benchmark to evaluate and compare the performance of several state-of-the-art metric learning based kinship verification methods. Experimental results are presented to demonstrate the efficacy of our proposed dataset and the effectiveness of existing metric learning methods for video-based kinship verification. Lastly, we also test human ability on kinship verification from facial videos and experimental results show that metric learning based computational methods are not as good as that of human observers.

* Corresponding author.
E-mail address: eyanhaibin@bupt.edu.cn (H. Yan).

http://dx.doi.org/10.1016/j.patcog.2017.03.001
0031-3203/

Table 1
Review and summary of existing representative kinship verification methods in the literature.

Method	Characteristics	Type	Year
Fang et al. [5]	Local feature representation	image	2010
Zhou et al. [6]	Local feature representation	image	2011
Xia et al. [7]	Transfer subspace learning	image	2012
Guo and Wang [9]	Bayes inference	image	2012
Zhou et al. [10]	Local feature representation	image	2012
Kohli et al. [18]	Local feature representation	image	2012
Somanath et al. [11]	Local feature representation	image	2012
Dibeklioglu et al. [19]	Dynamic feature representation	image	2013
Lu et al. [13]	Distance metric learning	image	2014
Guo et al. [14]	Logistic regression	image	2014
Yan et al. [15]	Multi-metric learning	image	2014
Yan et al. [20]	Mid-feature learning	image	2015
Our work	Distance metric learning	video	

The rest of this paper is organized as follows. In Section 2, we briefly review some related work, and Section 3 introduces the Kinship Face Videos in the Wild (KFVW) dataset. Section 4 presents some popular metric learning methods which have been widely used in kinship verification. Section 5 presents the experimental results and analysis. Finally, Section 6 concludes this paper.

2. Related work

In this section, we briefly review the related topics to our work: (1) kinship verification, (2) metric learning, and (3) video-based face analysis.

2.1. Kinship verification

The first study on kinship verification from facial images was made in [5]. In their work, they extracted local features such as skin color, gray value, histogram of gradient, and facial structure information in facial images and select some of them for kinship verification. Since this seminal work, more and more kinship verification methods have been proposed in the literature [5–7,9,10,13,15,18–23]. These methods can be mainly categorized into two classes: feature-based [5,6,9,10,19] and model-based [7,13–15]. Methods in the first class extract discriminative feature descriptors to represent kin-related information. Representative such feature information include skin color [5], histogram of gradient [5,6,11], Gabor wavelet [7,10,11,24], gradient orientation pyramid [10], local binary pattern [13,25], scale-invariant feature transform [11,13,15], salient part [8,9], self-similarity [18], and dynamic features combined with spatio-temporal appearance descriptor [19]. Methods in the second class learn discriminative models to verify kin relationship from face pairs. Typical such models are subspace learning [7], metric learning [13,15], transfer learning [7], multiple kernel learning [10] and graph-based fusion [14]. Table 1 lists a review and summary of existing representative kinship verification methods in the literature. All these kinship verification methods determine human kinship relationship from still face images, which may not discriminative enough to verify human kin relationship since large variations of human faces usually occur in still images.

2.2. Metric learning

A variety of metric learning methods [26–28,28–52] have been widely used in numerous computer vision tasks such as face recognition [26,28], gait recognition [34], object recognition, human activity recognition [29], human age estimation [30], person re-identification [28,31,32], visual tracking, and visual search. These methods can be mainly classified into two classes: unsupervised and supervised. The first class of methods learn a low-dimensional manifold to preserve the geometrical structure of data points, and the second class of methods seek an appropriate distance metric to exploit the discriminative information of samples. Recently, metric learning techniques have also been used in kinship verification [13,15], these methods are strongly supervised and require the exact label information of samples. For kinship verification, it is more convenient to obtain the weakly supervision of samples so that it is desirable to employ and evaluate weakly supervised methods for kinship verification.

Table 2
Compassion of existing facial datasets for kinship verification.

Dataset	Number of kinship pairs	Type	Year
CornellKin [5]	150	image	2010
UB KinFace [7]	400	image	2012
IIITD Kinship [18]	272	image	2012
Family101 [12]	206	image	2013
KinFaceW-I [13]	533	image	2014
KinFaceW-I [13]	1000	image	2014
TSKinFace [66]	2030	image	2015
KFVW (Ours)	418	video	

2.3. Video-based face recognition

A variety of video-based face analysis methods have been proposed in the literature, and these methods can be mainly classified into parametric [53–56] and nonparametric [57–65] methods. Parametric methods represent each face video as a parametric family of probabilistic distribution, and use the Kullback–Leibler divergence to measure the similarity of two face videos. However, these methods usually fail when the underlying distributional assumptions do not hold for different face videos. Nonparametric methods exploit geometrical information to measure the similarity of two face videos by modeling face each video as a single linear subspace or as the union of linear subspaces. While a variety of video-based face recognition methods have been presented, there is no work on video-based kinship verification, probably due to the lack of such datasets. In this work, we fill this gap and contribute a video dataset for kinship verification.

3. The kinship face videos in the wild dataset

In past few years, several facial datasets have been released to advance the kinship verification problem, e.g., CornellKin [5], UB KinFace [7], IIITD Kinship [18], Family101 [12], KinFaceW-I [13], KinFaceW-II [13], TSKinFace [66], etc. Table 2 provides a summary of existing facial datasets for kinship verification. However, these datasets only consist of still face images, in which each subject usually has a single face image. Due to the large variations of human faces, a single still image may not be discriminative enough to verify human kin relationship. To address these shortcomings, we collected a new video face dataset called Kinship Face Videos in the Wild (KFVW) for the video-based kinship verification study. Compared to a still image, a face video provides more information to describe the appearance of human face, because it can easily capture the face of the person of interest from different poses, expressions, and illuminations.

The KFVW dataset was collected from TV shows on the Web. We totally collected 418 pairs of face videos, and each video contains about 100 – 500 frames with large variations such as pose, lighting, background, occlusion, expression, makeup, age, etc. The average size of a video frame is about 900 × 500 pixels. There are four kinship relation types in the KFVW dataset: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S), and Mother-Daughter (M-D), and there are 107, 101, 100, and 110 pairs of kinship face videos for kin relationships F-S, F-D, M-S, and M-D respectively.

Fig. 1. Sampled video frames of our KFVW dataset. Each row lists three face images of a video. From top to bottom are Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D) kin relationships, respectively.

Fig. 1 shows several examples of our KFVW dataset for each kinship relations. We can see that the KFVW dataset depicts faces of the person of interest from different poses, expressions, background, and illuminations such that it can provide more information to describe the appearance of human face.

4. Video-based kinship verification using metric learning

Metric learning involves seeking a suitable distance metric from a training set of data points. Following the evaluation and settings used in Ref. [67], we employ several distance metric learning methods as baseline methods for the video-based kinship verification problem. These metric learning methods include information theoretic metric learning (ITML) [68], side-information based linear discriminant analysis (SILD) [69], KISS metric learning (KISSME) [28], and cosine similarity metric learning (CSML) [70]. This section briefly presents these metric learning methods.

4.1. ITML

Let $\mathbf{X} = [x_1, x_2, \cdots, x_N] \in \mathbb{R}^{d\times N}$ be a training set consisting of N data points, the aim of common distance metric learning approaches is to seek a positive semi-definite (PSD) matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ under which the squared Mahalanobis distance of two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ can be computed by:

$$d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j), \tag{1}$$

where d is the dimension of data point $\mathbf{x}_i$.

Information-theoretic metric learning (ITML) [68] is a typical metric learning method, which exploits the relationship of the multivariate Gaussian distribution and the set of Mahalanobis distances to generalize the regular Euclidean distance. The basic idea of ITML method is to find a PSD matrix $\mathbf{M}$ to approach a predefined matrix $\mathbf{M}_0$ by minimizing the LogDet divergence between two matrices under the constraints that the squared distance $d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j)$ of a positive pair (or similar pair) is smaller than a positive threshold τ_p while that of a negative pair (or dissimilar pair) is larger than a threshold τ_n, and we have $\tau_n > \tau_p > 0$. By employing this constraint on all pairs of training set, ITML can be formulated as the following LogDet optimization problem:

$$\begin{aligned} \min_{\mathbf{M}} \; & D_{ld}(\mathbf{M}, \mathbf{M}_0) = \mathrm{tr}(\mathbf{M}\mathbf{M}_0^{-1}) - \log\det(\mathbf{M}\mathbf{M}_0^{-1}) - d \\ \text{s.t.} \; & d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \le \tau_p \qquad \forall\ \ell_{ij} = 1 \\ & d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \ge \tau_n \qquad \forall\ \ell_{ij} = -1, \end{aligned} \tag{2}$$

in which the predefined metric $\mathbf{M}_0$ is set to the identity matrix in our experiments, $\mathrm{tr}(\mathbf{A})$ is the *trace* operation of a square matrix $\mathbf{A}$, and ℓ_{ij} means the pairwise label of a pair of data points $\mathbf{x}_i$ and $\mathbf{x}_j$, which is labeled as $\ell_{ij} = 1$ for a similar pair (with kinship) and $\ell_{ij} = -1$ for a dissimilar pair (without kinship). In practice, to solve the optimization problem (2), iterative Bregman projections are employed to project the present solution onto a single constraint by the following scheme:

$$\mathbf{M}_{t+1} = \mathbf{M}_t + \beta\ \mathbf{M}_t(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}_t, \tag{3}$$

in which β is a projection variable which is controlled by both the learning rate and the pairwise label of a pair of data points.

4.2. SILD

Side-information based linear discriminant analysis (SILD) [69] makes use of the side-information of pairs of data points to estimate the within-class scatter matrix $\mathbf{C}_p$ by employing positive pairs and the between-class scatter matrix $\mathbf{C}_n$ by using negative pairs in training set:

$$\mathbf{C}_p = \sum\nolimits_{\ell_{ij}=1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \tag{4}$$

$$\mathbf{C}_n = \sum\nolimits_{\ell_{ij}=-1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T. \tag{5}$$

Then, SILD learns a discriminative linear projection $\mathbf{W} \in \mathbb{R}^{d\times m}$, $m \le d$ by solving the optimization problem:

$$\max_{\mathbf{W}} \ \frac{\det(\mathbf{W}^T\mathbf{C}_n\mathbf{W})}{\det(\mathbf{W}^T\mathbf{C}_p\mathbf{W})}. \tag{6}$$

By diagonalizing $\mathbf{C}_p$ and $\mathbf{C}_n$ as:

$$\mathbf{C}_p = \mathbf{U}\mathbf{D}_p\mathbf{U}^T, \qquad (\mathbf{U}\mathbf{D}_p^{-1/2})^T\mathbf{C}_p(\mathbf{U}\mathbf{D}_p^{-1/2}) = \mathbf{I}, \tag{7}$$

$$(\mathbf{U}\mathbf{D}_p^{-1/2})^T\mathbf{C}_n(\mathbf{U}\mathbf{D}_p^{-1/2}) = \mathbf{V}\mathbf{D}_n\mathbf{V}^T, \tag{8}$$

the projection matrix $\mathbf{W}$ can be computed as:

$$\mathbf{W} = \mathbf{U}\mathbf{D}_p^{-1/2}\mathbf{V}, \tag{9}$$

in which matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, and matrices $\mathbf{D}_p$ and $\mathbf{D}_n$ are diagonal. In the transformed subspace, the squared Euclidean distance of a pair of data points $\mathbf{x}_i$ and $\mathbf{x}_j$ is calculated by:

$$\begin{aligned} d^2_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) &= \left\| \mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j \right\|_2^2 \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W}\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j). \end{aligned} \tag{10}$$

This distance is equivalent to computing the squared Mahalanobis distance in the original space, and we have $\mathbf{M} = \mathbf{W}\mathbf{W}^T$.

4.3. KISSME

Keep it simple and straightforward (KISS) metric learning (KISSME) [28] method aims to learn a distance metric from the perspective of statistical inference. KISSME makes the statistical decision whether a pair of data points $\mathbf{x}_i$ and $\mathbf{x}_j$ is dissimilar/negative or not by using the scheme of likelihood ratio test. The hypothesis $\mathcal{H}_0$ states that a pair of data points is dissimilar, and the hypothesis $\mathcal{H}_1$ states that this pair is similar. The log-likelihood ratio is shown as:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log\left(\frac{p(\mathbf{x}_i, \mathbf{x}_j|\mathcal{H}_0)}{p(\mathbf{x}_i, \mathbf{x}_j|\mathcal{H}_1)}\right), \tag{11}$$

where $p(\mathbf{x}_i, \mathbf{x}_j|\mathcal{H}_0)$ is the probability distribution function of a pair of data points under the hypothesis $\mathcal{H}_0$. The hypothesis $\mathcal{H}_0$ is accepted if $\delta(\mathbf{x}_i, \mathbf{x}_j)$ is larger than a nonnegative constant, otherwise the hypothesis $\mathcal{H}_0$ is rejected and this pair is similar. By assuming the single Gaussian distribution of the pairwise difference $\mathbf{z}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ and relaxing the problem (11), $\delta(\mathbf{x}_i, \mathbf{x}_j)$ is simplified as:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{C}_p^{-1} - \mathbf{C}_n^{-1})(\mathbf{x}_i - \mathbf{x}_j), \tag{12}$$

in which the covariance matrices $\mathbf{C}_p$ and $\mathbf{C}_n$ are computed by Eqs. (4) and (5) respectively.

To achieve the PSD Mahalanobis matrix $\mathbf{M}$, KISSME projects $\hat{\mathbf{M}} = \mathbf{C}_p^{-1} - \mathbf{C}_n^{-1}$ onto the cone of the positive semi-definite matrix $\mathbf{M}$ by clipping the spectrum of $\hat{\mathbf{M}}$ via the scheme of eigenvalue decomposition.

4.4. CSML

Unlike above three metric learning methods, cosine similarity metric learning (CSML) [70] method hopes to achieve a transformation $\mathbf{W} \in \mathbb{R}^{d\times m}$ with $m \leq d$ to compute cosine similarity of a pair of data points in the transformed subspace:

$$\begin{aligned} cs_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) &= \frac{(\mathbf{W}^T\mathbf{x}_i)^T(\mathbf{W}^T\mathbf{x}_j)}{\|\mathbf{W}^T\mathbf{x}_i\| \ \|\mathbf{W}^T\mathbf{x}_j\|} \\ &= \frac{\mathbf{x}_i{}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\sqrt{\mathbf{x}_i{}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i}\sqrt{\mathbf{x}_j{}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}}. \end{aligned} \tag{13}$$

To obtain $\mathbf{W}$, CSML minimizes the cross-validation error and formulates the following objective function:

$$\begin{aligned} \max_{\mathbf{W}} \ F(\mathbf{W}) &= \sum_{\ell_{ij}=1} cs_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - \alpha \sum_{\ell_{ij}=-1} cs_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) - \beta\|\mathbf{W} - \mathbf{W}_0\|^2, \end{aligned} \tag{14}$$

in which $\mathbf{W}_0$ is a prior matrix, the nonnegative constant α weights the contributions of positive pairs and negative pairs to the margin, and β balances the tradeoff between the regularization term $\|\mathbf{W} - \mathbf{W}_0\|^2$ and margin. Last, the gradient-based scheme is employed to find the solution $\mathbf{W}$. Ref. [70] provides more details of the optimization on solving CSML method.

Fig. 2. Cropped face images of our KFVW dataset. Each row lists three face images of a video. From top to bottom are Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D) kin relationships, respectively.

5. Experiments

In this section, we evaluated several state-of-the-art metric learning methods for video-based kinship verification on the KFVW dataset, and provided some baseline results on this dataset.

5.1. Experimental settings

For a video, we first detected face region of interest in each frame using a popular face detector described in [71], and then resized and cropped each face region into the size of 64 × 64 pixels. Table 2 shows the detected faces of several videos. In our experiments, if the number of frames of a video is more than 100, we just randomly detected 100 frames of this video. All cropped face images were converted to gray-scale, and we extracted the local binary patterns (LBP) [72] on these images. For each cropped face image of a video, we divided each image into 8 × 8 non-

Table 3
The EER (%) and AUC (%) of several metric learning methods using LBP feature on the KFVW dataset.

Method	Measure	F-S	F-D	M-S	M-D	Mean
Euclidean	EER	43.81	48.10	43.50	44.09	44.87
	AUC	60.49	56.02	57.83	58.91	58.31
ITML	EER	42.86	44.29	40.50	**42.73**	42.59
	AUC	59.11	56.79	61.50	**63.08**	60.12
SILD	EER	42.86	**42.86**	43.00	44.09	43.20
	AUC	62.64	**60.71**	58.47	59.04	60.21
KISSME	EER	40.00	44.76	43.50	42.73	42.75
	AUC	63.68	60.06	57.08	58.56	59.85
CSML	EER	**38.57**	47.14	**38.50**	43.18	**41.85**
	AUC	**66.23**	57.11	**64.36**	59.62	**61.83**

Table 4
The EER (%) and AUC (%) of several metric learning methods using HOG feature on the KFVW dataset.

Method	Measure	F-S	F-D	M-S	M-D	Mean
Euclidean	EER	47.14	47.62	45.00	42.73	45.62
	AUC	56.44	54.85	54.84	59.30	56.36
ITML	EER	47.14	48.10	45.00	**41.82**	45.51
	AUC	55.98	54.09	57.09	59.08	56.56
SILD	EER	43.33	**43.81**	**42.00**	43.18	**43.08**
	AUC	**59.66**	57.04	59.68	59.74	**59.03**
KISSME	EER	44.76	44.29	43.00	45.91	44.49
	AUC	58.39	**57.85**	**61.04**	56.77	58.51
CSML	EER	**42.86**	47.62	45.00	44.09	44.89
	AUC	59.51	56.07	59.76	**59.79**	58.78

overlapping blocks, in which the size of each block is 8 × 8 pixels, and then we extracted a 59-bin uniform pattern LBP histogram for each block and concatenated histograms of all blocks to form a 3776-dimensional feature vector. To obtain the feature representation for each cropped face video, we averaged the feature vectors of all frames within this video to form a mean feature vector in this benchmark. Then, principal component analysis (PCA) was employed to reduce dimensionality of each vector to 100 dimension.

In this benchmark, we used all positive pairs for each kinship relation, and also generated the same number of negative pairs. The positive pair (or true pair) means that there is a kinship relation between a pair of face videos. The negative pair (or false pair) denotes that there is not a kinship relation between a pair of face videos. Specifically, a negative pair consists of two videos, one was randomly selected from the parents' set, and another who is not his/her true child was randomly selected children's set. For each kinship relation, we randomly took 80% of video pairs for model training and the rest 20% pairs for testing. We repeated this procedure 10 times, and recorded the Receiver Operating Characteristic (ROC) curve for performance evaluation, under which two measures: the Equal Error Rate (EER) and the Area Under an ROC Curve (AUC) were adopted to report the performance of various metric learning methods for video-based kinship verification. Note that small EER and large AUC show high performance of a method.

5.2. Results and analysis

This subsection presents the results and analysis of different methods on KFVW dataset for video-based kinship verification.

5.2.1. Comparisons of different metric learning methods

We first evaluated several metric learning methods using LBP features for video-based kinship verification, and provided the baseline results on the KFVW dataset. The baseline methods include Euclidean, ITML [68], SILD [69], KISSME [28], and CSML [70]. The Euclidean method means that the similarity/dissimilarity between a pair of face videos is computed by Euclidean distance in the original space. The metric learning method first learns a distance metric from the training data itself, and then employs this learned distance metric to calculate the distance of a pair of videos from the testing data. Table 3 shows the EER (%) and AUC (%) of these metric learning methods using LBP feature on the KFVW dataset. From this table, we see that (1) CSML obtains the best performance in terms of the mean EER and mean AUC, and also achieves the best EER and AUC on the F-S and M-S subsets; (2) ITML shows the best performance on the M-D subset; (3) SILD obtains the best EER and AUC on the F-D subset; (4) all metric learning based methods, i.e., ITML, SILD, KISSME and CSML, outperform Euclidean method in terms of the EER and AUC; (5) most of methods achieve the best performance on F-S subset compared with other three subsets; and (6) the best EER is merely about 38.5%, and thus video-based kinship verification on the KFVW dataset is extremely challenging. Moreover, Fig. 3 plots ROC curves of several metric learning methods using LBP feature on the KFVW dataset for four types of kinship relations.

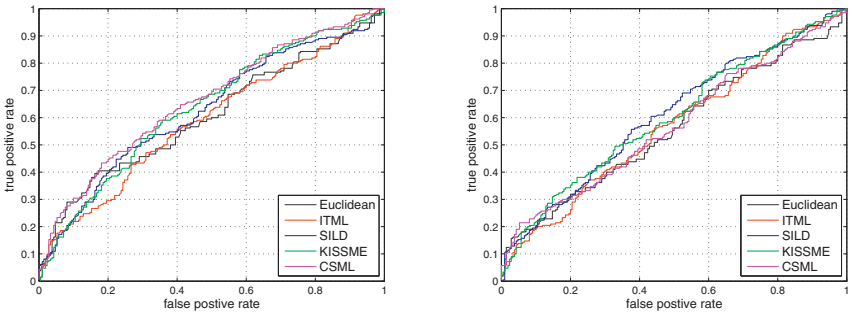
5.2.2. Comparisons of different feature descriptors

We also evaluated several state-of-the-art metric learning methods using different feature descriptors. To this end, we extracted the histogram of oriented gradients (HOG) [73] from two different scales for each cropped face image. Specifically, we first divided each image into 16 × 16 non-overlapping blocks, where the size of each block is 4 × 4 pixels. Then, we divided each image 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 HOGs of all blocks to form a 2880-dimensional feature vector. Following the same procedure as in extracting LBP, for a cropped face video, we averaged the feature vectors of all frames within this video to yield a mean feature vector as the final feature representation. Then, PCA was employed to reduce dimensionality of each vector to 100 dimension.

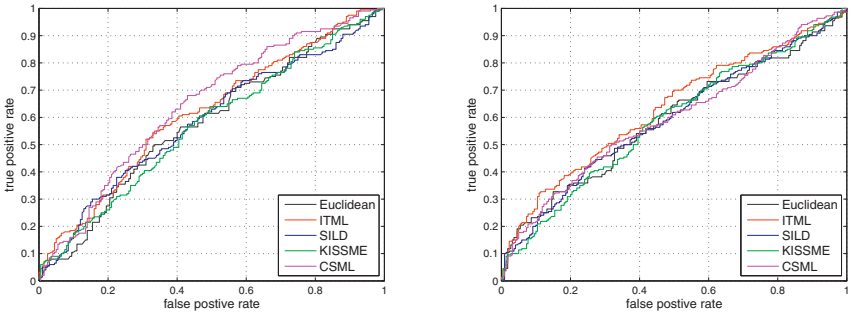
Table 4 reports the EER (%) and AUC (%) of several metric learning methods using HOG feature on the KFVW dataset, and Fig. 4 shows ROC curves of these methods using HOG feature. From this table, we see that 1) SILD achieves the best performance in terms of the mean EER and mean AUC, and also obtains the best EER on the F-D and M-S subsets; and 2) KISSME obtains the best AUC on the F-D and M-S subsets. By comparing Tables 3 and 4, we see that metric learning methods using LBP feature outperform the same methods using HOG feature in terms of the mean EER and mean AUC. The reason may be that LBP feature can capture local texture characteristics of face images which is more useful than gradient characteristics extracted by HOG feature to help improve the performance of video-based kinship verification.

5.2.3. Parameter analysis

We investigated how different dimensions of LBP feature affect the performance of these state-of-the-art metric learning methods. Figs. 5–8 show the EER and the AUC (%) of ITML, SILD, KISSME, and CSML methods versus different dimensions of LBP feature on the KFVW dataset for four types of kin relationships, respectively. From these figures, we see that (1) ITML and CSML methods show the relatively stable AUC on four subsets (i.e., F-S, F-D, M-S, and M-D) by increasing the dimension of LBP feature from 10 to 100; and (2) SILD and KISSME methods achieve the best AUC at dimension of 30 and then gradually reduce AUC with the increasing of dimension of LBP feature from 30 to 100. Therefore, we reported the EER and the AUC of these metric learning methods at dimension of 30 on four subsets for fair comparison.

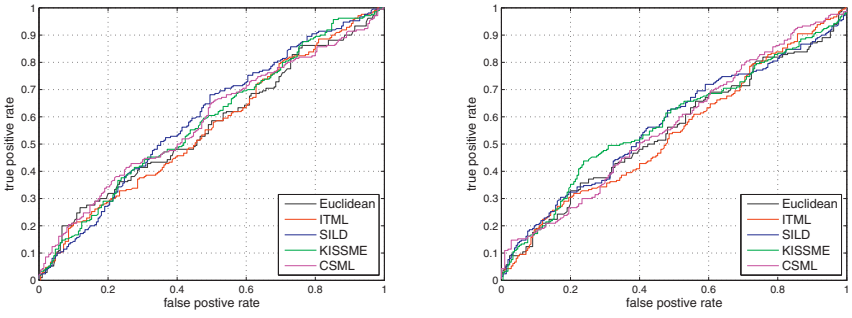


(a) F-S (b) F-D

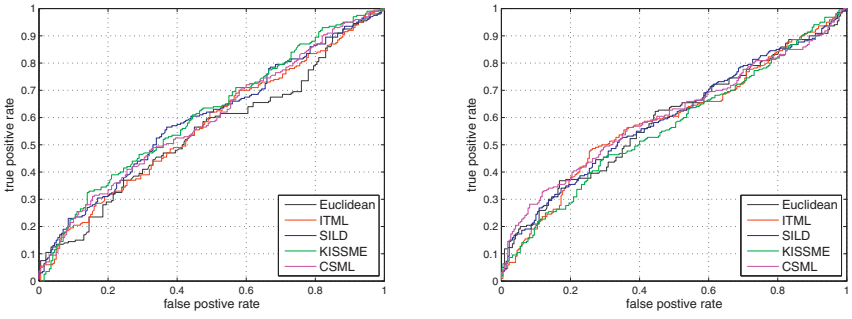


(c) M-S (d) M-D

Fig. 3. ROC curves of several metric learning methods using LBP feature on our KFVW dataset for four types of kinship relations.

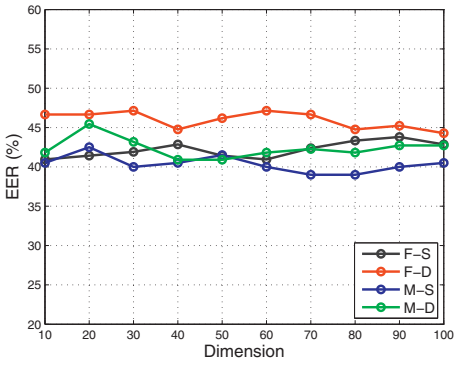


(a) F-S (b) F-D

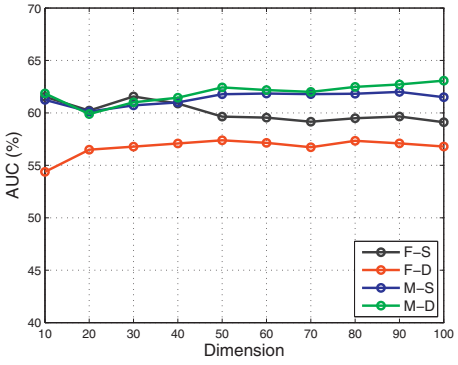


(c) M-S (d) M-D

Fig. 4. ROC curves of several metric learning methods using HOG feature on our KFVW dataset for four types of kinship relations.

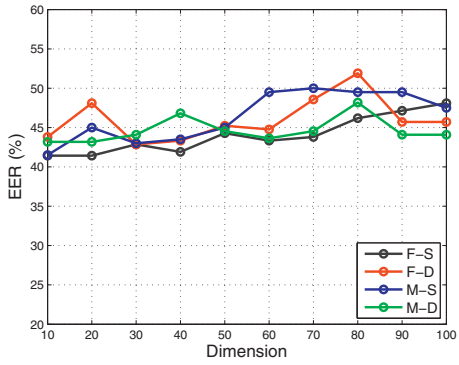


(a) EER

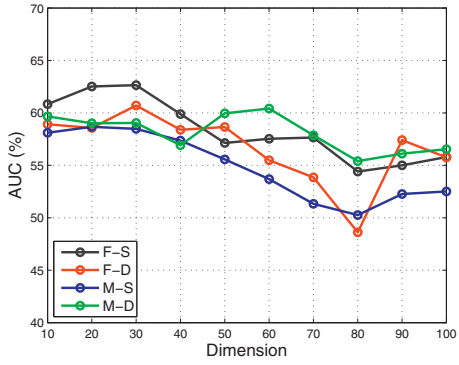


(b) AUC

Fig. 5. The EER and AUC (%) of ITML method using LBP feature on the KFVW dataset.



(a) EER



(b) AUC

Fig. 6. The EER and AUC (%) of SILD method using LBP feature on the KFVW dataset.

5.2.4. Computational cost

We conducted experiments on a standard Windows machine (Intel i5-3470 CPU @ 3.20 GHz, and 32GB RAM) with the MATLAB code. Given a face video, detecting face region of interest of a frame takes about 0.9 s, and extracting LBP feature of a cropped face image with size of 64 × 64 takes about 0.02 s. In model training, the training times of ITML, SILD, KISSME, and CSML methods are around 9.6, 0.6, 0.7, and 6.5 s for each kin relationship, respectively. In testing, the matching times of these methods are about 0.02 s (excluding times of face detection and feature extraction) for a pair of face videos.

5.2.5. Human observers for kinship verification

As another baseline, we also evaluated human ability to verify kin relationship from face videos on the KFVW dataset. For each kinship relation, we randomly chose 20 positive pairs of face videos and 20 negative pairs of face videos, and displayed these video pairs for ten volunteers to decide whether there is a kin relationship or not. These volunteers consist of five male students and five female students, whose ages range from 18 to 25 years, and they have not experienced any training on verifying kin relationship from face videos. We designed two tests (i.e., Test A and Test B) to examine the human ability to verify kin relationship from face videos. In Text A, the cropped face videos were provided to human volunteers, and volunteers did decision making on the detected face regions with size of 64 × 64 pixels. In Text B, the original face videos were presented to volunteers, and human volunteers can make their decisions by exploiting multiple cues in the whole images, e.g., skin color, hair, race, background, etc. Table 5 lists the mean verification accuracy (%) of human ability on video-based kinship verification for different types of kin relationships

Table 5
The mean verification accuracy (%) of human ability on video-based kinship verification on the KFVW dataset four types of kin relationships.

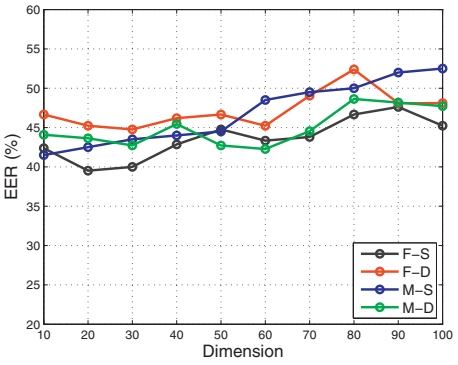
Method	F-S	F-D	M-S	M-D	Mean
Test A	70.50	66.50	67.50	70.00	68.63
Test B	75.00	70.50	73.00	73.50	73.00

on the KFVW dataset. We see that Test B reports better performance Test A on four kinship relations. The reason is that Test B can exploit more cues such as hair and background to help make correct make correct kinship verification. From this table, we also observe that human observers provide higher verification accuracy than metric learning-based methods on KFVW dataset.

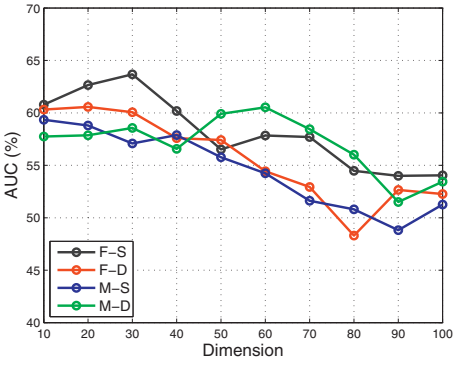
5.3. Discussions

From experimental results shown in Tables 3–5 and Figs. 3–8, we make the following observations:

- State-of-the-art metric learning methods outperform predefined metric-based method (i.e., Euclidean distance) for video-based kinship verification. The reason is that metric learning method can learn a distance metric from the training data itself to increase the similarity of a positive pair and to decrease the similarity of a negative pair in the learned metric space.
- LBP feature presents the better performance than HOG feature for video-based kinship verification. The reason may be that LBP feature can encode local texture characteristics of face images which is more useful than gradient characteristics ex-

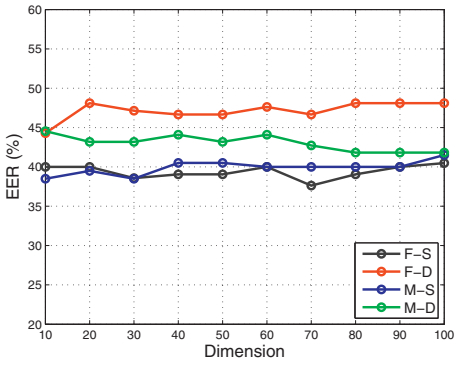


(a) EER

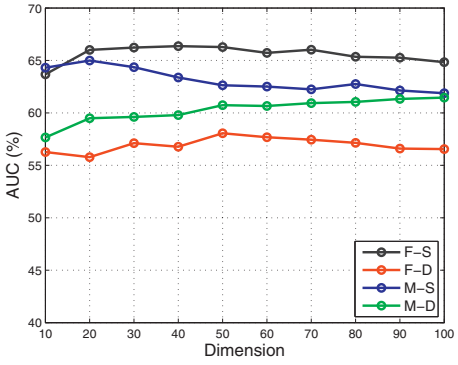


(b) AUC

Fig. 7. The EER and AUC (%) of KISSME method using LBP feature on the KFVW dataset.



(a) EER



(b) AUC

Fig. 8. The EER and AUC (%) of CSML method using LBP feature on the KFVW dataset.

tracted by HOG feature to help improve the performance of video-based kinship verification.

- Metric leaning methods and human observers achieve the poor performance on F-D subset compared with other three subsets, which shows that kinship verification on F-D subset is a more challenging task.
- The best EER of metric learning methods is merely about 38.5%, thus it is very challenging to advance the study of video-based kinship verification on the KFVW dataset.

6. Conclusion

In this paper, we have studied the problem of video-based kinship verification. To our best knowledge, this problem has not been formally addressed in the literature. We have first presented a new video face dataset called Kinship Face Videos in the Wild (KFVW) which were captured in wild conditions for the video-based kinship verification study, as well as the standard benchmark. Then, we have evaluated and compared the performance of several state-of-the-art metric learning based kinship verification methods. Lastly, we also have tested test human ability on kinship verification from facial videos and experimental results show that metric learning based computational methods are not as good as that of human observers. Experimental results are presented to demonstrate the efficacy of our proposed dataset and the effectiveness of existing metric learning methods for video-based kinship verification.

In our future work, we plan to design more efficient feature learning method for video face representation and advanced video-based distance metric learning to further improve the performance of video-based kinship verification.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61603048, the Beijing Natural Science Foundation under Grant 4174101, and the Fundamental Research Funds for the Central Universities.

References

[1] M.F. Dal Martello, L.T. Maloney, Where are kin recognition signals in the human face? J. Vision 6 (12) (2006) 1356–1366.
[2] A. Alvergne, R. Oda, C. Faurie, A. Matsumoto-Oda, V. Durand, M. Raymond, Cross-cultural perceptions of facial resemblance between kin, J. Vision 9 (6) (2009) 1–10.
[3] L.M. DeBruine, F.G. Smith, B.C. Jones, S. Craig Roberts, M. Petrie, T.D. Spector, Kin recognition signals in adult faces, Vision Res. 49 (1) (2009) 38–43.
[4] G. Kaminski, S. Dridi, C. Graff, E. Gentaz, Human ability to detect kinship in strangers' faces: effects of the degree of relatedness, Proc. R. Soc. B 276 (1670) (2009) 3193–3200.
[5] R. Fang, K.D. Tang, N. Snavely, T. Chen, Towards computational models of kinship verification, in: IEEE International Conference on Image Processing, 2010, pp. 1577–1580.
[6] X. Zhou, J. Hu, J. Lu, Y. Shang, Y. Guan, Kinship verification from facial images under uncontrolled conditions, in: ACM International Conference on Multimedia, 2011, pp. 953–956.
[7] S. Xia, M. Shao, J. Luo, Y. Fu, Understanding kin relationships in a photo, IEEE Trans. Multimedia 14 (4) (2012) 1046–1056.
[8] S. Xia, M. Shao, Y. Fu, Toward kinship verification using visual attributes, in: International Conference on Pattern Recognition, 2012, pp. 549–552.
[9] G. Guo, X. Wang, Kinship measurement on salient facial features, IEEE Trans. Instrum. Meas. 61 (8) (2012) 2322–2325.
[10] X. Zhou, J. Lu, J. Hu, Y. Shang, Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments, in: ACM International Conference on Multimedia, 2012, pp. 725–728.
[11] G. Somanath, C. Kambhamettu, Can faces verify blood-relations? in: IEEE International Conference on Biometrics: Theory, Applications and Systems, 2012, pp. 105–112.

[12] R. Fang, A.C. Gallagher, T. Chen, A. Loui, Kinship classification by modeling facial feature heredity, in: IEEE International Conference on Image Processing, 2013, pp. 2983–2987.
[13] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2) (2014) 331–345.
[14] Y. Guo, H. Dibeklioglu, L. van der Maaten, Graph-based kinship recognition, in: International Conference on Pattern Recognition, 2014, pp. 4287–4292.
[15] H. Yan, J. Lu, W. Deng, X. Zhou, Discriminative multimetric learning for kinship verification, IEEE Trans. Inf. Forensics Secur. 9 (7) (2014) 1169–1178.
[16] J. Lu, J. Hu, X. Zhou, J. Zhou, M.C. Santana, J. Lorenzo-Navarro, L. Kou, Y. Shang, A. Bottino, T.F. Vieira, Kinship verification in the wild: the first kinship verification competition, in: IEEE International Joint Conference on Biometrics, 2014, pp. 1–6.
[17] J. Lu, J. Hu, V.E. Liong, X. Zhou, A. Bottino, I.U. Islam, T.F. Vieira, X. Qin, X. Tan, S. Chen, S. Mahpod, Y. Keller, L. Zheng, K. Idrissi, C. Garcia, S. Duffner, A. Baskurt, M.C. Santana, J. Lorenzo-Navarro, The FG 2015 kinship verification in the wild evaluation, in: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 2015, pp. 1–7.
[18] N. Kohli, R. Singh, M. Vatsa, Self-similarity representation of weber faces for kinship classification, in: IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2012, pp. 245–250.
[19] H. Dibeklioglu, A.A. Salah, T. Gevers, Like father, like son: facial expression dynamics for kinship verification, in: IEEE International Conference on Computer Vision, 2013, pp. 1497–1504.
[20] H. Yan, J. Lu, X. Zhou, Prototype-based discriminative feature learning for kinship verification, IEEE Trans. Cybern. 45 (11) (2015) 2535–2545.
[21] S. Xia, M. Shao, Y. Fu, Kinship verification through transfer learning, in: International Joint Conference on Artificial Intelligence, 2011, pp. 2539–2544.
[22] M. Shao, S. Xia, Y. Fu, Genealogical face recognition based on ub kinface database, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2011, pp. 60–65.
[23] J. Hu, J. Lu, J. Yuan, Y.-P. Tan, Large margin multi-metric learning for face and kinship verification in the wild, in: Asian Conference on Computer Vision, 2014, pp. 252–267.
[24] S. Du, R.K. Ward, Improved face representation by nonuniform multilevel selection of Gabor convolution features, IEEE Trans. Syst. Man Cybern. Part B 39 (6) (2009) 1408–1419.
[25] W. Deng, J. Hu, J. Guo, Extended src: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.
[26] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? Metric learning approaches for face identification, in: IEEE International Conference on Computer Vision, 2009, pp. 498–505.
[27] J. Lu, Y.-P. Tan, Regularized locality preserving projections and its extensions for face recognition, IEEE Trans. Syst. Man Cybern. Part B 40 (3) (2010) 958–963.
[28] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2288–2295.
[29] D. Tran, A. Sorokin, Human activity recognition with metric learning, in: European Conference on Computer Vision, 2008, pp. 548–561.
[30] B. Xiao, X. Yang, Y. Xu, H. Zha, Learning distance metric for regression by semidefinite programming with application to human age estimation, in: ACM International Conference on Multimedia, 2009, pp. 451–460.
[31] W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by probabilistic relative distance comparison, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 649–656.
[32] A. Mignon, F. Jurie, Pcca: a new approach for distance learning from sparse pairwise constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2666–2672.
[33] J. Hu, J. Lu, Y.-P. Tan, Discriminative deep metric learning for face verification in the wild, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1875–1882.
[34] J. Lu, G. Wang, P. Moulin, Human identity and gender recognition from gait sequences with arbitrary walking directions, IEEE Trans. Inf. Forensics Secur. 9 (1) (2014) 51–61.
[35] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[36] J. Lu, Y.-P. Tan, Ordinary preserving manifold analysis for human age and head pose estimation, IEEE Trans. Hum. Mach. Syst. 43 (2) (2013) 249–258.
[37] J. Lu, Y.-P. Tan, Uncorrelated discriminant nearest feature line analysis for face recognition, IEEE Signal Process. Lett. 17 (2) (2010) 185–188.
[38] J. Lu, Y.-P. Tan, A doubly weighted approach for appearance-based subspace learning methods, IEEE Trans. Inf. Forensics Secur. 5 (1) (2010) 71–81.
[39] J. Lu, Y.-P. Tan, Cost-sensitive subspace learning for face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2661–2666.
[40] J. Lu, Y.-P. Tan, Nearest feature space analysis for classification, IEEE Signal Process. Lett. 18 (1) (2011) 55–58.
[41] J. Lu, E. Zhang, Gait recognition for human identification based on ica and fuzzy svm through multiple views fusion, Pattern Recognit. Lett. 28 (16) (2007) 2401–2411.
[42] J. Lu, V.E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2041–2056.
[43] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[44] J. Lu, V.E. Liong, J. Zhou, Cost-sensitive local binary feature learning for facial age estimation, IEEE Trans. Image Process. 24 (12) (2015) 5356–5368.
[45] J. Lu, V.E. Liong, G. Wang, P. Moulin, Joint feature learning for face recognition, IEEE Trans. Inf. Forensics Secur. 10 (7) (2015) 1371–1383.
[46] J. Lu, G. Wang, W. Deng, K. Jia, Reconstruction-based metric learning for unconstrained face verification, IEEE Trans. Inf. Forensics Secur. 10 (1) (2015) 79–89.
[47] J. Lu, Y.-P. Tan, Cost-sensitive subspace analysis and extensions for face recognition, IEEE Trans. Inf. Forensics Secur. 8 (3) (2013) 510–519.
[48] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Cost-sensitive semi-supervised discriminant analysis for face recognition, IEEE Trans. Inf. Forensics Secur. 7 (3) (2012) 944–953.
[49] J. Lu, V.E. Liong, J. Zhou, Simultaneous local binary feature learning and encoding for face recognition, in: 2015 IEEE International Conference on Computer Vision, 2015, pp. 3721–3729.
[50] J. Lu, G. Wang, W. Deng, P. Moulin, J. Zhou, Multi-manifold deep metric learning for image set classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1137–1145.
[51] V.E. Liong, J. Lu, G. Wang, P. Moulin, J. Zhou, Deep hashing for compact binary codes learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2475–2483.
[52] H. Liu, J. Lu, J. Feng, J. Zhou, Group-aware deep feature learning for facial age estimation, Pattern Recognit. (2016, accepted).
[53] G. Shakhnarovich, J. Fisher, T. Darrell, Face recognition from long-term observations, in: European Conference on Computer Vision, 2006, pp. 361–375.
[54] K.C. Lee, J. Ho, M.H. Yang, D. Kriegman, Video-based face recognition using probabilistic appearance manifolds, in: IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. 313–320.
[55] A. Hadid, M. Pietikainen, From still image to video-based face recognition: an experimental analysis, in: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 2004, pp. 813–818.
[56] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, T. Darrell, Face recognition with image sets using manifold density divergence, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 581–588.
[57] T.K. Kim, J. Kittler, R. Cipolla, Discriminative learning and recognition of image set classes using canonical correlations, IEEE Trans. Pattern Anal. Mach. Intell. 29 (6) (2007) 1005–1018.
[58] R. Wang, S. Shan, X. Chen, W. Gao, Manifold-manifold distance with application to face recognition based on image set, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[59] R. Wang, X. Chen, Manifold discriminant analysis, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1–8.
[60] H. Cevikalp, B. Triggs, Face recognition based on image sets, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2567–2573.
[61] M.T. Harandi, C. Sanderson, S. Shirazi, B.C. Lovell, Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 2705–2712.
[62] Y. Hu, A.S. Mian, R. Owens, Sparse approximated nearest points for image set classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 121–128.
[63] Y. Hu, A.S. Mian, R. Owens, Face recognition using sparse approximated nearest points between image sets, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1992–2004.
[64] K. Fan, W. Liu, S. An, X. Chen, Margin preserving projection for image set based face recognition, in: International Conference on Neural Information Processing, 2011, pp. 681–689.
[65] J. Lu, G. Wang, P. Moulin, Localized multifeature metric learning for image-set-based face recognition, IEEE Trans. Circuits Syst. Video Technol. 26 (3) (2016) 529–540.
[66] X. Qin, X. Tan, S. Chen, Tri-subject kinship verification: understanding the core of a family, IEEE Trans. Multimedia 17 (10) (2015) 1855–1867.
[67] J. Hu, J. Lu, Y.-P. Tan, Fine-grained face verification: dataset and baseline results, in: International Conference on Biometrics, 2015, pp. 79–84.
[68] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: International Conference on Machine Learning, 2007, pp. 209–216.
[69] M. Kan, S. Shan, D. Xu, X. Chen, Side-information based linear discriminant analysis for face recognition, in: British Machine Vision Conference, 2011, pp. 1–12.
[70] H.V. Nguyen, L. Bai, Cosine similarity metric learning for face verification, in: Asian Conference on Computer Vision, 2010, pp. 709–720.
[71] M. Mathias, R. Benenson, M. Pedersoli, L.J.V. Gool, Face detection without bells and whistles, in: European Conference on Computer Vision, 2014, pp. 720–735.
[72] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[73] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

Haibin Yan received the B.E. and M.E. degrees from the Xi'an University of Technology, Xi'an, China, in 2004 and 2007, and the Ph.D. degree from the National University of Singapore, Singapore, in 2013, all in mechanical engineering. Now, she is an Assistant Professor in the School of Automation, Beijing University of Posts and Telecommunications, Beijing, China. From October 2013 to July 2015, she was a research fellow at the Department of Mechanical Engineering, National University of Singapore, Singapore. Her research interests include service robotics and computer vision.

Junlin Hu received the B.E. degree from the Xian University of Technology, Xian, China, in 2008, and the M.E. degree from the Beijing Normal University, Beijing, China, in 2012. He is currently pursuing the Ph.D. degree with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His current research interests include computer vision, pattern recognition, and biometrics.