ORIGINAL ARTICLE

A weighted exponential discriminant analysis through side-information for face and kinship verification using statistical binarized image features

Oualid Laiadi[1,4] · Abdelmalik Ouamane[2] · Abdelhamid Benakcha[3] · Abdelmalik Taleb-Ahmed[4] · Abdenour Hadid[4]

Abstract

Side-information based exponential discriminant analysis (SIEDA) is more efficient than side-information based linear discriminant analysis (SILDA) in computing the discriminant vectors because it maximizes the Fisher criterion function. In this paper, we develop a novel criterion, named side-information based weighted exponential discriminant analysis (SIWEDA), that is based on the classical SIEDA method. We reformulate and generalize the classical Fisher criterion function in order to maximize it, with the property to pull as close as possible the intra-class samples (within-class samples), and push and repulse away as far as possible the inter-class samples (between-class samples). Thus, SIWEDA selects the eigenvalues of high significance and eliminate those with less discriminative information. To reduce the feature vector dimensionality and lighten the class intra-variability, we use SIWEDA and within class covariance normalization (WCCN) using the proposed statistical binarized image features (StatBIF). Moreover, we use score fusion strategy to extract the complementarity of different weighting scales of our StatBIF descriptor. We conducted experiments to evaluate the performance of the proposed method under unconstrained environment, using five datasets namely LFW, YTF, Cornell KinFace, UB KinFace and TSKin-Face datasets, in the context of matching faces and kinship verification in the wild conditions. The experiments showed that the proposed approach outperforms the current state of the art. Very interestingly, our approach showed superior performance compared to methods based on deep metric learning.

Keywords Kinship verification · Face matching · Unconstrained environment · Weighting factor · SIWEDA · StatBIF · Fisher criterion

1 Introduction

Biometric facial images include large human traits, such as identity, gender, expression, age, and ethnicity [5, 36, 44]. Over the past two decades, automatic face recognition under controlled conditions showed increasing results and best performances in wide-scale biometric topics. However, satisfactory performances are still beyond reach because of the big challenge of weakly labeled data setup and under unconstrained environments, in which the latter is characterized by blurring images, low quality, without any restrictions in terms of pose, background, expression, lighting, and partial occlusion. More recently, many public datasets [16, 24, 49, 57, 61] studied face and kinship applications that deal with facial images in unconstrained environments. Most of the applications deal with a scenario of verifying a pair of facial images to check whether the images belong to the same person or to different persons. Therefrom, check the same person who appears in the passport photo is a typical example where identity verification via face can operate, especially by the security organizations [7, 18, 37].

Therefore, pairs of facial images are used for the training stage, over weakly labeled data that denotes if the pair of images belong to the same person or the two facial images belong to two different persons. Whereas in test stage, a new pair of facial images is presented to check the appropriate

✉ Oualid Laiadi
  oualid.laiadi@gmail.com

1 Laboratory of LESIA, University of Biskra, Biskra, Algeria

2 University of Biskra, Biskra, Algeria

3 Laboratory of LGEB, University of Biskra, Biskra, Algeria

4 Univ. Polytechnique Hauts-de-France, CNRS, Univ. Lille, ISEN, Centrale Lille, UMR 8520 - IEMN - DOAE, F-59313 Valenciennes, France

decision if it is matched/mismatched to the person [23]. Other face application under unconstrained environments through weakly labeled data is checking if two facial images belong to persons from the same family or not. The challenge is to understand and extract the face similarities between family members, because many research works have investigated and proved that facial traits are significant cues to verify the kinship [16, 33, 34, 49, 63].

To tackle this problem, several algorithms have been proposed in the previous literature works to increase the performance of the verification systems. Subspace transformation methods are mostly used for face recognition, which is considered as one of the efficient applied methods in this topic, the mutual role of using these methods is predominantly to transform the high dimensional features into a new lower and discriminative subspace. Which means, enhancing the separation between the classes [66]. The most used and well-known subspace transformation methods are principle component analysis (PCA) [8] and linear discriminant analysis (LDA) [46].

Unsupervised subspace transformation methods cannot properly conserve the features structure model of different classes (i.e., the structure of each data class is not taken into consideration). Discriminative features are often conserved by supervised subspace transformation methods. LDA (also called Fisher's linear discriminant or FLD) is the traditional method which learns extracted features and transforms them into discriminant subspace. Unfortunately, it cannot be directly applied because of the small size sample (SSS) problem [3, 50] at the reason of the within-class scatter matrix's singularity. As we know, face recognition is a typical small size problem. Many works have been reported to use LDA for face recognition. The mostly used approach, called Fisherface (or PCA+LDA), was proposed by Belhumeur et al. [6] and Swets et al. [53]. In their approach, PCA was first employed to minimize the size of the original feature space from M to d, and then the classical FLD is applied to minimize the dimension from d to n ($n \leq d$). Whereas some useful discriminative information may be lost, caused by the features that are thrown away in the PCA phase. On the other hand, the PCA phase cannot guarantee the success of transforming the within-class scatter matrix to be nonsingular.

Later on, the exponential discriminant analysis (EDA) [66] method is proposed to solve the singularity problem, which has showed interesting results in face recognition, by projecting the features through nonlinear subspace. However, the supervised subspace dimensionality reduction and transformation methods like LDA and EDA need the full labeled class information. Sometimes this information is not available for all classes under the unconstrained settings (i.e., the only information is that pair is match or mismatch/kin or nonkin). Therefore, Kan et al. [38],

proposed subspace transformation method called Side-information based Linear Discriminant Analysis (SILD), in which they proposed a substitutional solution to compute the between-class scatter matrix (S_b) and within-class scatter matrix (S_w). Besides, defining a new representation that directly can be computed through the side-information (i.e., weakly supervised information, where only the pairwise label information is available to train the methods). Inevitably, SILD method suffers from the SSS problem, thus the authors suggested the solution of Fisherface approach applied in LDA-based methods. Recently, Ouamane et al. [43], proposed an effective method named side-information based exponential discriminant analysis (SIEDA). Inspired by EDA, the authors benefit from the advantages of the matrix exponential which preserved the discriminative information included in the null space of the within-class scatter matrix unlike to PCA+SILD approach. In addition, a kernel method (diffusion) is used for the transformation of the nonlinear posed problems across linear problems. Similarly, to the kernel methods, the within-class and between-class scatter matrices are transformed through a new space by the exponential function [32, 45, 56, 59, 64, 65]. Motivated by this research, we propose in this paper a novel Side-information based weighted exponential discriminant analysis (SIWEDA) method for face and kinship verification in the wild. We reformulate the Fisher criterion of the classical SIEDA to maximize the distance of the between-class scatter matrix and minimize the distance of the within-class scatter matrix using a weighting factor (α) in order to make best separation between classes. Moreover, in unconstrained environment, the applicability of the texture information may be limited by the image degradations. For this reason, feature extraction is a key step that promoted and incubated from the facial recognition research community to describe facial images (i.e. Local Binary Patterns (LBP) [41], Local Phase Quantization (LPQ) [42], Binarized Statistical Image Features (BSIF) [25], Context-Aware Local Binary Feature Learning (CA-LBFL) [14], Rotation-Invariant Local Binary Descriptor (RI-LBD) [13] and Deep Binary Descriptor with Multi-Quantization (DBD-MQ) [15]). We propose new local face descriptor called Statistical Binarized Image Features (StatBIF). Our descriptor shows more efficient traits unlike the well-known local descriptors in the literature works.

The purpose of this paper is to investigate the efficiency of the proposed framework based on local histogram features of the local descriptors, Multi-scale Local Binary Patterns (MLBP), Multi-scale Local Phase Quantization (MLPQ), Multi-scale Binarized Statistical Image Features (MBSIF), and Statistical Binarized Image Features (StatBIF). The contributions of this work can be summarized as follows:

1. We introduce a novel local feature for describing facial images. Our descriptor is based on the local statistics traits of the facial image and the original BSIF operator.
2. We propose a new method SIWEDA for face and kinship verification based on the classical SIEDA method. Furthermore, to lighten the class intra-variability, we proposed two variants SIEDA+WCCN and SIWEDA+WCCN by integrating WCCN in SIEDA and SIWEDA, respectively.
3. We extensively evaluate our approach against the state-of-the-art methods using five challenging face and kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, YTF and LFW databases.

The rest of the paper is organized as follows: Sect. 2 gives demonstration of the proposed descriptor for face and kinship verification. We present and describe our SIWEDA method in Sect. 3. The experimental data and setup are presented and results are discussed in Sect. 4. Finally, concluding remarks are given in Sect. 5.

2 Statistical binarized image features (StatBIF)

We propose a new operator to describe kin relation facial images inspired by the statistical operator and the Binarized Statistical Image Features (BSIF) operator [25]. The BSIF operator 1 computes an image patch code I of size $l \times l$ pixels and a linear filter Z_i which contains N filters of the same size, the filter response s_i is obtained by

$$s_i = \sum_{m,n} Z_i(m, n) I(m, n) \tag{1}$$

In order to enhance the discrimination of BSIF feature to local changes in the facial image and to increase its robustness, we determine a new image space described by the local statistical features extraction. Considering the same method as in LBP operator, each pixel of a given facial image is represented by a statistical value calculated using its neighborhood pixels to form a new image. Thus defined by the number of pixels P and the circle of radius R. When the new statistical images are produced, the basic BSIF operator, with filter size $l \times l$, is applied to each image separately. The binarized feature b_i is obtained by setting $b_i = 1$ if $s_i > 0$ and $b_i = 0$ otherwise. Formally, the Statistical BIF feature can be obtained by decimal conversion of the binarization of the filters response:

$$StatBIF = \sum_{i=1}^{N} b_i (\sum_{m,n} Z_i(m, n) Stat_{P,R}(m, n)) \times 2^{i-1} \tag{2}$$

where P and R are defined as the circle for calculating the statistical value of the pixel where ($R = \frac{l-1}{2}$ and $P = R \times N$) and $l \times l$ is the filter size for computing the BSIF code (see Fig. 1). In Eq. (2), Stat refers to the statistical function used to generate the new local statistical representation. In our experiments, we consider six statistical functions:

– The median:

$$median_{P,R}(i) = median\{i_{p=0}, \dots, i_{p=P-1}\} \tag{3}$$

– The mean:

$$mean_{P,R}(i) = \frac{1}{P} \sum_{p=0}^{P-1} i_p \tag{4}$$

– The standard deviation:

$$std_{P,R}(i) = \sqrt{\frac{\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2}{P-1}} \tag{5}$$

– The variance:

$$variance_{P,R}(i) = \frac{1}{P} \sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2 \tag{6}$$

– The skewness:

$$skewness_{P,R}(i) = \frac{\frac{1}{P} \sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^3}{\left(\sqrt{\frac{1}{P} \sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2}\right)^{3/2}} \tag{7}$$
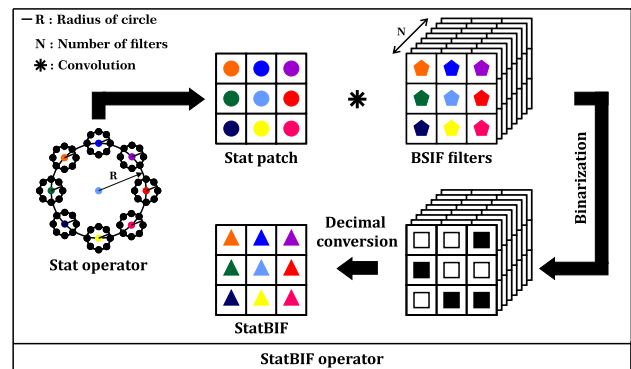
– The kurtosis:

Fig. 1 Computation of the statistical binarized image features. First, the local statistics of the image are estimated on the circle (P, R). Then, the original BSIF operator, with parameters $l \times l$, is applied

$$kurtosis_{P,R}(i) = \frac{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^4}{\left(\frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2\right)^2} \qquad (8)$$

The computation of StatBIF features is illustrated in Fig. 1. Examples of images obtained by applying StatBIF feature for different statistical functions, on facial image are shown in Fig. 2.

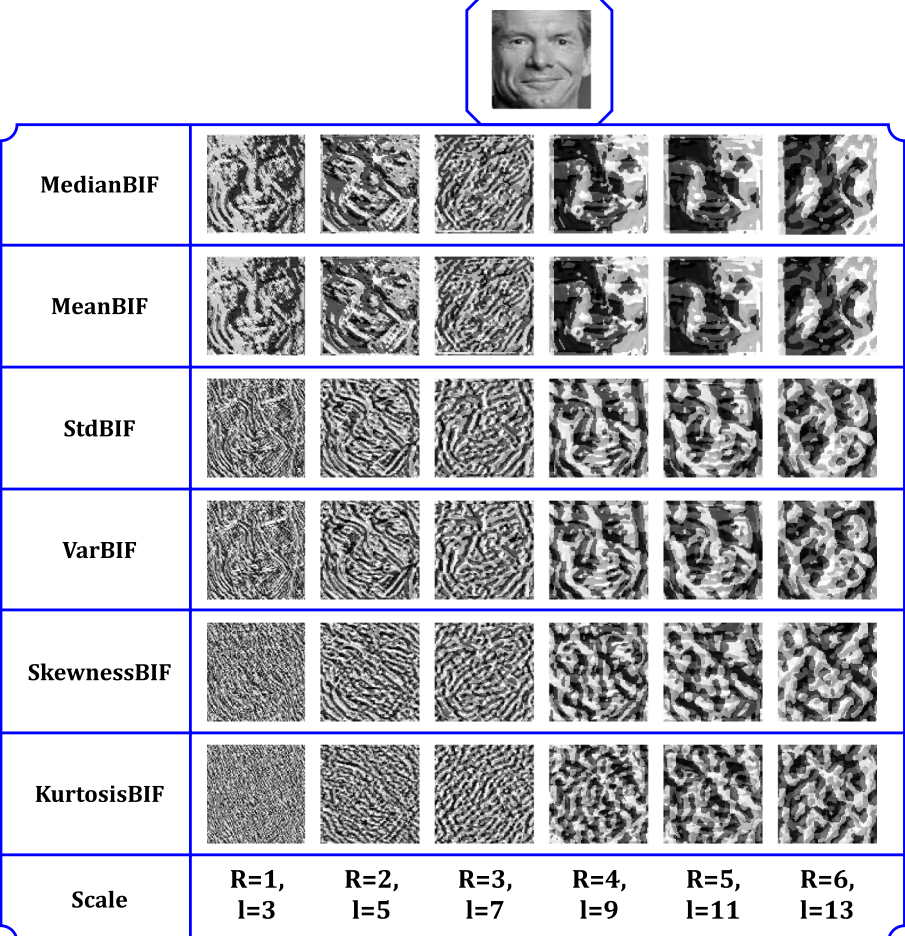## 3 Side-information based weighted exponential discriminant analysis (SIWEDA)
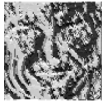
3.1 Matrix exponential
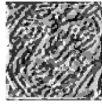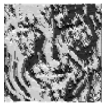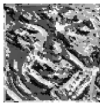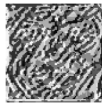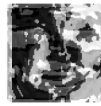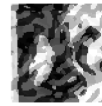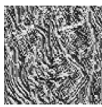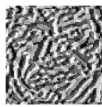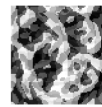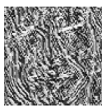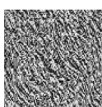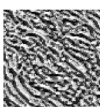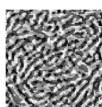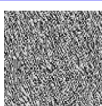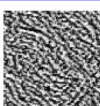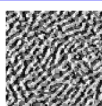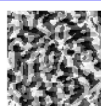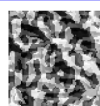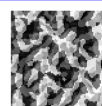
To introduce SIWEDA method we are going to give the following definitions and properties. Let X be a $q \times q$ square matrix with its exponential denoted by $exp(X)$ or e^X, is defined as:

$$exp(X) = \sum_{j=0}^{\infty} \frac{X^j}{j!} = I + X + \frac{X^2}{2!} + \cdots + \frac{X^n}{n!} + \cdots,$$ where I is the unit matrix with the same size of X. Following are some properties of the matrix exponential:

1. $exp(0) = I$.
2. $exp(X)$ is a full rank matrix.
3. If $XY = YX$, then $exp(X + Y) = exp(X)exp(Y) = exp(Y)exp(X)$.
4. $(exp(X))^{-1} = exp(-X)$.
5. For the invertible matrix Z, $exp(Z^{-1}XZ) = Z^{-1}exp(X)Z$.
6. If $X = diag(x_1, x_2, \ldots, x_m)$ is a diagonal matrix, $exp(X) = diag(exp(x_1), exp(x_2), \ldots, exp(x_m))$.
7. $|exp(X)| = exp(tr(X))$.
8. If $(v_1, v_2, \ldots, v_m)$ are eigenvectors for matrix X and $\Lambda_1, \Lambda_2, \ldots, \Lambda_m$ the corresponding eigenvalues, then $(v_1, v_2, \ldots, v_m)$ are also eigenvectors of exp(X) that correspond to the eigenvalues $exp(\Lambda_1), exp(\Lambda_2), \ldots, exp(\Lambda_m)$.

Fig. 2 Examples of the generated faces by applying the statistical binarized image features, with various scales

MedianBIF						
MeanBIF						
StdBIF						
VarBIF						
SkewnessBIF						
KurtosisBIF						
Scale	R=1, l=3	R=2, l=5	R=3, l=7	R=4, l=9	R=5, l=11	R=6, l=13

3.2 Side-information based exponential discriminant analysis (SIEDA)

Supervised subspace transformation methods, such as linear discriminant analysis (LDA)[46] and exponential discriminant analysis (EDA) [66], enhance the discrimination of the extracted features by transforming these data in a sub-space where it is easier to perform its separation and classification. These methods need the class information for each sample. The within class scatter matrix (S_w) and the between class scatter matrix (S_b) must be calculated with full label information:

$$S_w = \sum_{i=1}^{L} \frac{1}{n_i} \sum_{j=1}^{n_i} (\xi_j^i - m^i)(\xi_j^i - m^i)^T \tag{9}$$

$$S_b = \sum_{i=1}^{L} (m^i - \bar{m})(m^i - \bar{m})^T \tag{10}$$

where L represents the total number of classes. The mean of the ith class is m^i and the total mean is $\bar{m}$. While ξ_j^i represents each sample included in the ith class, and n_i represents the number of samples contained in the ith class.

That means, LDA and EDA fail in weakly labeled data. Kan et al. [38] proposed a new representation to resolve this problem by directly operating the S_w and S_b matrices with the side-information. The positive classes pair images are directly utilized to calculate the within class scatter matrix and the negative classes pair images are used to compute the between class scatter matrix. Let us refer $P_{class} = \left\{ (\breve{\xi}_i^1, \hat{\xi}_i^1) : l(\breve{\xi}_i^1) = l(\hat{\xi}_i^1) \right\}$ as the collection of positive-class image pairs and $N_{class} = \left\{ (\breve{\xi}_i^0, \hat{\xi}_i^0) : l(\breve{\xi}_i^0) \neq l(\hat{\xi}_i^0) \right\}$ as the collection of negative-class image pairs, where the image ξ is represented by the class label $l(\xi)$. Here, the within-class and between-class scatter matrices of side-information based linear discriminant analysis (SILD) method can be represented by:

$$S_w^{sild} = \sum_{i=1}^{C_1} (\breve{\xi}_i^1 - \hat{\xi}_i^1)(\breve{\xi}_i^1 - \hat{\xi}_i^1)^T \tag{11}$$

$$S_b^{sild} = \sum_{i=1}^{C_0} (\breve{\xi}_i^0 - \hat{\xi}_i^0)(\breve{\xi}_i^0 - \hat{\xi}_i^0)^T \tag{12}$$

The target function for SILD is:

$$\begin{aligned} U_{opt}^{sild} &= argmax_U \frac{U^T S_b^{sild} U}{U^T S_w^{sild} U} \\ &= argmax_U \frac{\left| U^T (V_b^T \Lambda_b V_b) U \right|}{\left| U^T (V_w^T \Lambda_w V_w) U \right|} \end{aligned} \tag{13}$$

where:

$$\begin{aligned} S_w^{sild} &= V_w^T \Lambda_w V_w \\ S_b^{sild} &= V_b^T \Lambda_b V_b \end{aligned} \tag{14}$$

To make better separation of the discriminative information, Ouamane *et al.* [43] proposed SIEDA method. Like EDA [66], SIEDA replaces the eigenvalues Λ_{w_k} by $exp(\Lambda_{w_k})$ in S_w^{sild}, and the eigenvalues Λ_{b_k} by $exp(\Lambda_{b_k})$ in S_b^{sild}. Thus, from property (8) of the matrix exponential the target function for SILD becomes:

$$\begin{aligned} U_{opt}^{sieda} &= argmax_U \frac{\left| U^T (V_b^T exp(\Lambda_b) V_b) U \right|}{\left| U^T (V_w^T exp(\Lambda_w) V_w) U \right|} \\ &= argmax_U \frac{U^T exp(S_b^{sild}) U}{U^T exp(S_w^{sild}) U} \end{aligned} \tag{15}$$

This problem is reduced to a generalized eigenvalue problem:

$$(exp(S_w^{sild}))^{-1} exp(S_b^{sild}) = V^T \Lambda V \tag{16}$$

The SIEDA transformation matrix is given by the eigenvectors V_k of $(exp(S_w^{sild}))^{-1} exp(S_b^{sild})$, ordered according to their corresponding eigenvalues Λ_k in descending order of magnitude.

In Eq. (13), the inter-class distance of training samples from different subjects $(\breve{\xi}_i^0, \hat{\xi}_i^0)|_{i=0}^{C_0}$ for all pairs are maximized by the numerator while the intra-class distance of training samples from the same subjects $(\breve{\xi}_i^1, \hat{\xi}_i^1)|_{i=0}^{C_1}$ for all pairs are minimized by the denominator. This equation derived from the following multiobjective programming problem:

$$\begin{aligned} &max \frac{U^T S_b^{sild} U}{U^T U} \\ &min \frac{U^T S_w^{sild} U}{U^T U} \end{aligned} \tag{17}$$

The between-class distance ($dist_b$) and the within-class distance ($dist_w$) can be calculated by trace of two scatter matrices: $dist_b = trace(S_b^{sild}) = \Lambda_{b_1} + \Lambda_{b_2} + \cdots + \Lambda_{b_n}$ and $dist_w = trace(S_w^{sild}) = \Lambda_{w_1} + \Lambda_{w_2} + \cdots + \Lambda_{w_n}$.

Whereas, from property (6) of the matrix exponential, the two distances becomes:

$dist_b = trace(exp(S_b^{sild})) = exp(\Lambda_{b_1}) + exp(\Lambda_{b_2}) + \cdots + exp(\Lambda_{b_n})$ and

$dist_w = trace(exp(S_w^{sild})) = exp(\Lambda_{w_1}) + exp(\Lambda_{w_2}) + \cdots + exp(\Lambda_{w_n})$.

Therefore, the ratio $exp(\Lambda_{b_k})/exp(\Lambda_{w_k})$ is bigger than $\Lambda_{b_k}/\Lambda_{w_k}$. Thus, we can conclude that there are differences in diffusion scale between the within and between-class distances and that leads to a best separation.

3.3 Proposed side-information based weighted exponential discriminant analysis (SIWEDA)

To enhance the separation between the positive and negative classes, we propose SIWEDA method to maximize the objective function of SIEDA method. ($\alpha = 1$) is the particular weighting factor value of the objective target function (15) of the classical SIEDA. We generalize the objective target function to contain and accept different values of weighting factor (α).

From (13) the generalized objective target function of SILD is:

$$U_{opt}^{siwld} = argmax_U \frac{U^T \alpha S_b^{sild} U}{U^T \alpha S_w^{sild} U}$$
$$= argmax_U \frac{\left| U^T \alpha(V_b^T \Lambda_b V_b) U \right|}{\left| U^T \alpha(V_w^T \Lambda_w V_w) U \right|} \quad (18)$$
$$= argmax_U \frac{\left| U^T (V_b^T \alpha \Lambda_b V_b) U \right|}{\left| U^T (V_w^T \alpha \Lambda_w V_w) U \right|}$$

where $\alpha > 0$.

Then, from property (8) of the matrix exponential the generalized target function becomes:

$$U_{opt}^{siweda} = argmax_U \frac{\left| U^T (V_b^T exp(\alpha \Lambda_b) V_b) U \right|}{\left| U^T (V_w^T exp(\alpha \Lambda_w) V_w) U \right|}$$
$$= argmax_U \frac{U^T exp(\alpha S_b^{sild}) U}{U^T exp(\alpha S_w^{sild}) U} \quad (19)$$

The aim of SIWEDA is to seek m discriminant vectors such the trace of between-class scatter matrix is maximized and the trace of within-class matrix is minimized. While the objective function for discriminant SIEDA criterion is to increase simultaneously the between-class distance and decrease the within-class distance. However, the difference between the two distances is limited by using the specific weighting factor ($\alpha = 1$).

By applying the weighting factor (α), the distance becomes:

$dist_b = trace(exp(\alpha S_b^{sild})) = exp(\alpha \Lambda_{b_1}) + exp(\alpha \Lambda_{b_2}) + \cdots + exp(\alpha \Lambda_{b_n})$ and

$dist_w = trace(exp(\alpha S_w^{sild})) = exp(\alpha \Lambda_{w_1}) + exp(\alpha \Lambda_{w_2}) + \cdots + exp(\alpha \Lambda_{w_n})$

Therefore, the ratio:

$$exp(\alpha \Lambda_{b_k})/exp(\alpha \Lambda_{w_k}) > exp(\Lambda_{b_k})/exp(\Lambda_{w_k}) > \Lambda_{b_k}/\Lambda_{w_k}$$
$$(20)$$

where $\alpha > 1$.

The projection matrix U_{opt}^{siweda} is composed by the most significant eigenvectors of $(exp(\alpha S_w^{sild}))^{-1} exp(\alpha S_b^{sild})$.

From SIWEDA method, we can deduce that there is a large difference in the diffusion scale across the within and between-class distances which enhance the separation.

For more explanation of our method, Fig. 3 illustrates an example of the proportion of Λ_k, $exp(\Lambda_k)$ and two weighting exponentials $exp(2 \times \Lambda_k)$ and $exp(5 \times \Lambda_k)$ and their respective sums. In this figure, the largest eigenvalue $\Lambda_5 = 33.33\%$ while its corresponding exponential $exp(\Lambda_5) = 63.64\%$, $exp(2 \times \Lambda_5) = 88.47\%$ and $exp(5 \times \Lambda_5) = 99.33\%$ and the smallest eigenvalue $\Lambda_1 = 6.67\%$ while its corresponding exponential $exp(\Lambda_1) = 1.17\%$, $exp(2 \times \Lambda_1) = 0.029\%$ and $exp(5 \times \Lambda_1) = 10^{-7}\%$. We have: $\Lambda_{b_k}/\Lambda_{w_k} < exp(\Lambda_{b_k})/exp(\Lambda_{w_k}) < exp(2 \times \Lambda_{b_k})/exp(2 \times \Lambda_{w_k}) < exp(5 \times \Lambda_{b_k})/exp(5 \times \Lambda_{w_k})$. We see that ($\alpha = 5$) neglected the Λ_4 which may contain a significant information with sum of 0.67%, unlike to ($\alpha = 2$) which conserve the Λ_4 with sum of 11.70%. This means that, each set of training data from different datasets is adapted with a specific weighting factor (α), that can be simplified the separation between the positive and negative classes.

In Sect. 4, our experiments show that by changing the value of the weighting factor (α) gives us the ability for easier selection the eigenvalues of high significance and eliminate those with less discriminative information. Thus, the small eigenvalues are reduced and large eigenvalues are enlarged.
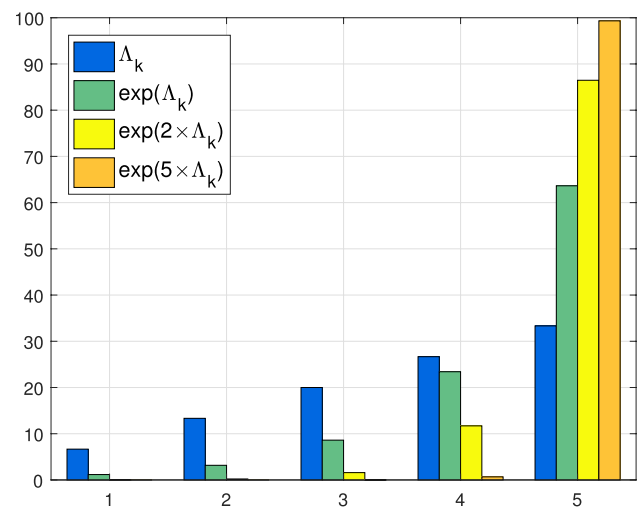
Fig. 3 Example of the proportions $\frac{\Lambda_k}{\sum \Lambda_k}$ (blue bars), $\frac{exp(\Lambda_k)}{\sum exp(\Lambda_k)}$ (green bars), $\frac{exp(2 \times \Lambda_k)}{\sum exp(2 \times \Lambda_k)}$ (yellow bars) and $\frac{exp(5 \times \Lambda_k)}{\sum exp(5 \times \Lambda_k)}$ (orange bars) (color figure online)

3.4 Within-class covariance normalization

The first use of the within-class covariance normalization (WCCN) is in the community of speaker recognition. While Dehak et al. [12] founded that it is the best technique to project the reduced-vectors of LDA method to a new subspace determined by the square-root of the inverse of the within-class covariance matrix. We propose a new variant of SIWEDA by integrating WCCN:

$$W = \sum_{i=1}^{C_1} \frac{(U^{siweda})^T \check{\xi}_i^1 - (U^{siweda})^T \hat{\xi}_i^1}{(U^{siweda})^T \check{\xi}_i^1 - (U^{siweda})^T \hat{\xi}_i^1} \tag{21}$$

where, U^{siweda} is the SIWEDA projection matrix found in Eq. (19). The WCCN projection matrix C is obtained by Cholesky decomposition of the inverse of W: $W^{-1} = CC^T$. Where the new projection matrix Z^{siweda} is obtained by: $Z^{siweda} = C^T U^{siweda}$. By imposing upper bounds on the classification error metric [4], WCCN decreases the within-class variations effect by reducing the expected classification error on the training step.

The procedure of this proposed variant, side-information weighted exponential discriminant analysis integrating within class covariance normalization (SIWEDA+WCCN), is detailed in algorithm 1.

Algorithm 1 Side-Information Weighted Exponential Discriminant Analysis plus Within Class Covariance Normalization

Input:
- The matrix ξ of the N training samples.
- The weak labels ($labels_W$) for extracting the positive-class image pairs
$P_{class} = \left\{ (\check{\xi}_i^1, \hat{\xi}_i^1) : l(\check{\xi}_i^1) = l(\hat{\xi}_i^1) \right\}$ and negative-class image pairs
$N_{class} = \left\{ (\check{\xi}_i^0, \hat{\xi}_i^0) : l(\check{\xi}_i^0) \neq l(\hat{\xi}_i^0) \right\}$
- α is the maximum separation weighting value.
Output:
- The projection matrix Z^{siweda} of SIWEDA.
Algorithm:

1: $S_w^{sild} = \sum_{i=1}^{C_1} (\check{\xi}_i^1 - \hat{\xi}_i^1)(\check{\xi}_i^1 - \hat{\xi}_i^1)^T$
2: $S_b^{sild} = \sum_{i=1}^{C_0} (\check{\xi}_i^0 - \hat{\xi}_i^0)(\check{\xi}_i^0 - \hat{\xi}_i^0)^T$
3: Compute the weighted matrices: $exp(\alpha S_w^{sild})$ and $exp(\alpha S_b^{sild})$
4: Compute the eigenvectors V_k and corresponding eigenvalues Λ_k of $(exp(\alpha S_w^{sild}))^{-1} exp(\alpha S_b^{sild})$.
5: Sort the m eigenvectors $U^{siweda} = V_k$ according to Λ_k in decreasing order.
6: $W = \sum_{i=1}^{C_1} ((U^{siweda})^T \check{\xi}_i^1 - (U^{siweda})^T \hat{\xi}_i^1)((U^{siweda})^T \check{\xi}_i^1 - (U^{siweda})^T \hat{\xi}_i^1)^T$
7: Compute WCCN projection matrix (C): $W^{-1} = CC^T$
8: Compute the new $Z^{siweda} = C^T U^{siweda}$

3.5 Similarity measure

We compute the similarity of the two feature vectors by the cosine distance in the SIWEDA+WCCN subspace as follows:

$$Cos(\check{\xi}, \hat{\xi}) = \frac{(C^T U^{siweda} \check{\xi})^T (C^T U^{siweda} \hat{\xi})}{\left\| (C^T U^{siweda} \check{\xi}) \right\| \left\| (C^T U^{siweda} \hat{\xi}) \right\|} \tag{22}$$

After discriminant analysis methods, the using of cosine similarity distance has an advantage comes from its connection to the Bayes decision rule, as the optimal used method is the Bayes classifier for decreasing the classification error [31].

4 Experiments

The experiments are organized into three parts: First part presents the benchmark datasets utilized in our experiments; Second part gives the parameter settings utilized in our framework; Third part provides the results with their discuss and compare the best ones with those of the state of the art.

4.1 Benchmark datasets

To value the effectiveness of the proposed kinship verification framework, the experiments are conducted on five well-known datasets: LFW dataset, YTF dataset, Cornell KinFace

dataset, UB KinFace v2 dataset, TSKinFace dataset. The facial images are with large age difference and various ethnicities, captured under uncontrolled environments and no restriction in terms of pose. In the following, we give a brief description of the five datasets.

Labeled faces in the wild (LFW) dataset [24] is a big dataset collected from the web, specially gathered to study the problem of face recognition in unconstrained environments containing real-world variations in terms of lighting, pose, expressions, blur, occlusion, resolution, and so on. This challenging dataset consists 13,233 facial images belonging to 5,749 different subjects.

YouTube face (YTF) dataset [57] consists of 3425 videos from 1595 different subjects with various variations of pose, expression and illumination, and the average length of each video clip is 181.3 frames.

Cornell KinFace dataset [16] contains 143 pairs of parents and children images collected from the Internet. In total, there are 286 cropped frontal facial images. The majority of the images were taken from Google Images. To make sure that the facial extracted characteristics are of high quality, only frontal facial images with a neutral facial expression are chosen. We note that, for privacy issue, seven families are taken out of the original dataset that consists of 150 families.

UB KinFace dataset [61] includes 600 images of 400 people, distributed as 200 pairs of child-young parent (set 1) and 200 pairs of child-old parent (set 2). Images in the dataset are of public figures (politicians and celebrities) from Internet.

TSKinFace dataset [49] contains two kinds tri-subject kinship relations: father-mother-daughter (FM-D) and father-mother-son (FM-S). The number of FM-D and the FM-S relations comprise 502 and 513, respectively (4060 facial images). These images are for public figures gathered from the Internet. For fair comparison, we restructured the dataset by separating the group of Father-Mother-Daughter into two groups Father-Daughter and Mother-Daughter kinship relations, and the group of Father-Mother-Son into two groups Father-Son and Mother-Son kinship relations. Therefore, we are restricted to the four kin relations as in the other datasets.

4.2 Parameter settings

Face matching

The LFW dataset is divided into two views: view 1 is utilized for model selection, and view 2 is put to evaluate performance, containing three evaluation paradigms: (1) the image unrestricted protocol, (2) the image restricted protocol and (3) the unsupervised protocol. In our experiments, the aligned images (LFW-a) were used, and the proposed approach was evaluated on the view 2 using image restricted protocol, where no outside additional training data

was used. The dataset is subdivided into ten disjoint folds cross-validation. While for training step, nine folds are used and the remaining fold for testing step. Each fold includes 300 matched (positive) pairs and 300 mismatched (negative) pairs. The final performance is reported as the mean accuracy $\pm$ standard deviation (SE) as well as the ROC curve through the tenfold cross-validation.

The YouTube Face (YTF) dataset we followed the same evaluation protocol founded in [57] of unconstrained face verification including 5000 video pairs, which were divided into 10 folds and each fold contains 250 positive pairs and 250 negative pairs. We learned the feature representation using MLBP, MLPQ and MBSIF descriptors and our Stat-BIF descriptor for each frame of video clips, separately. As all facial images have been aligned already, we averaged all descriptors of one video clip to make a mean vector as the feature of the video.

Kinship verification

We evaluated the performance of the proposed framework on the same experimental protocol cited in the literature works [34, 62], whither fivefold cross-validation for kinship verification is carried out, while preserving the number of pairs images nearly equal in all folds. By following this protocol, we ensure that our results directly compared to the state of the art. We generated the negative pairs of the kinship randomly such that the appearance of every facial image is only once in the training set. In the training and test phases, the number of negative pairs and positive pairs is equal.

Features extraction

Concerning the face normalization, all face attributes descriptions are extracted from images that are aligned and cropped into a resolution of 64×64 pixels. We extract at multi-scale the local binary patterns (MLBP) [41], the radius $R = \{1, 2, 3, 4, 5, 6\}$ and the number of pixels in the neighborhood $P = 8$. For multi-scale binarized statistical image features (MBSIF) [25] we use eight filters with different sizes $W = \{3, 5, 7, 9, 11, 13\}$. In the multi-scale local phase quantization (MLPQ) [42], the window size is $M = \{3, 5, 7, 9, 11, 13\}$. For Statistical Binarized Image Features (Stat-BIF), we use six scales, $StatBIF_{l=3}$, $StatBIF_{l=5}$, $StatBIF_{l=7}$, $StatBIF_{l=9}$, $StatBIF_{l=11}$ and $StatBIF_{l=13}$. From each scale the extracted features are the combination of six statistical features of (median, mean, standard deviation, variance, skewness and kurtosis).

Every facial image is subdivided into 36 blocks, each of size 12×12 pixels. By using histograms of 256 bins, we assemble the local features extracted from each block. Thereafter, we concatenate the histograms of all the 36 blocks, where the dimension of the obtained vector from each descriptor is $6 \times 36 \times 256$.

Table 1 Mean verification accuracy (%) of our SIEDA+WCCN method compared to the classical SIEDA method using StatBIF, MLBP, MLPQ and MBSIF descriptors on Cornell KinFace, UB KinFace, TSKinFace and LFW

Descriptor	Cornell	UB KinFace			TSKinFace					LFW
	Mean	Set 1	Set 2	Mean	F-S	F-D	M-S	M-D	Mean	Mean ± std
MLBP+SIEDA	71.39	71.51	67.46	69.49	80.02	79.48	81.19	81.06	80.44	93.97 ± 1.01
MLPQ+SIEDA	73.85	72.23	68.69	70.46	81.88	80.38	83.43	83.55	82.31	93.67 ± 0.97
MBSIF+SIEDA	74.94	72.01	68.66	70.34	80.71	80.18	81.48	82.95	81.33	94.43 ± 0.96
MLBP+SIEDA+WCCN	73.83	72.51	69.44	70.98	80.80	79.97	82.26	82.06	81.27	93.97 ± 1.01
MLPQ+SIEDA+WCCN	75.25	72.22	70.91	71.57	82.47	80.78	83.72	84.65	82.91	94.30 ± 0.86
MBSIF+SIEDA+WCCN	75.95	72.51	70.90	71.71	82.07	81.38	82.65	84.35	82.61	**94.43 ± 0.96**
$StatBIF_{l=3}$ + SIEDA	76.89	72.73	69.95	71.34	81.10	80.98	81.28	83.75	81.78	92.20 ± 1.17
$StatBIF_{l=5}$ + SIEDA	73.12	72.73	70.66	71.70	81.19	82.57	82.16	84.75	82.67	93.50 ± 0.97
$StatBIF_{l=7}$ + SIEDA	73.56	72.74	69.41	71.08	80.90	81.08	82.94	83.85	82.19	93.60 ± 1.07
$StatBIF_{l=9}$ + SIEDA	74.61	70.76	68.91	69.84	80.90	80.08	82.84	84.05	81.97	93.77 ± 1.01
$StatBIF_{l=11}$ + SIEDA	75.62	71.26	68.67	69.97	81.68	79.68	82.65	82.86	81.72	93.57 ± 0.99
$StatBIF_{l=13}$ + SIEDA	76.28	70.72	68.91	69.82	81.19	78.59	81.77	82.86	81.10	94.13 ± 0.92
$StatBIF_{l=3}$ + SIEDA + WCCN	76.91	73.22	71.42	72.32	82.37	81.87	82.16	84.75	82.79	92.60 ± 1.12
$StatBIF_{l=5}$ + SIEDA + WCCN	75.91	73.23	**71.66**	**72.44**	82.46	**82.87**	83.04	**86.05**	**83.61**	93.87 ± 0.93
$StatBIF_{l=7}$ + SIEDA + WCCN	75.28	**73.74**	71.13	72.44	81.97	81.37	**83.82**	85.15	83.08	94.00 ± 1.00
$StatBIF_{l=9}$ + SIEDA + WCCN	75.63	71.74	70.14	70.94	81.97	80.58	83.62	84.25	82.61	94.20 ± 0.92
$StatBIF_{l=11}$ + SIEDA + WCCN	77.01	72.24	69.65	70.95	**82.94**	81.37	83.53	83.56	82.85	94.03 ± 0.96
$StatBIF_{l=13}$ + SIEDA + WCCN	**77.71**	72.73	70.41	71.57	82.85	79.69	83.43	83.86	82.46	94.27 ± 0.90

The numbers in bold represent the best results for each column

4.3 Results and discussion

To verify our framework, we generate several experiments listed in Tables 1 and 2. The mean verification accuracies of our SIEDA+WCCN method compared to SIEDA using the proposed StatBIF descriptor and the three existing descriptors MLBP, MLPQ and MBSIF on the experimental datasets is shown in Table 1. As can be seen in this table, the best performance for kinship verification using StatBIF descriptor through SIEDA+WCCN method outperforms the use of this descriptor through SIEDA method only. Whereas, our StatBIF descriptor shows the best accuracy with scales, $l = 13$ for Cornell KinFace dataset and $l = 5$ for both, UB KinFace and TSKinFace datasets. As we can see from LFW dataset, SIEDA+WCCN failed to project the StatBIF features into discriminative subspace, because there is a limitation of separation over the specific weighting factor ($\alpha = 1$).

Table 2 demonstrates the mean verification accuracy of our SIWEDA+WCCN method using different descriptors with different weighting factors on the experimental datasets. This table shows that the weighting factors ($\alpha > 1$) have the best performance compared with the SIEDA+WCCN method ($\alpha = 1$). Also, there is specific value of (α) that was adapted to the training set of different datasets, which increase the performance of face and kinship verification through the SIWEDA method. We see that by using SIWEDA+WCCN method over different weighting descriptors ($StatBIF_{l=11,\alpha=5}$), ($StatBIF_{l=5,\alpha=20}$), ($MBSIF_{\alpha=15}$), ($StatBIF_{l=9,\alpha=15}$), and ($StatBIF_{l=9,\alpha=15}$) show the best accuracies compared to the classical SIEDA+WCCN method ($\alpha = 1$) for Cornell KinFace dataset, UB KinFace dataset, TSKinFace dataset, LFW dataset and YTF dataset, respectively.

4.4 Weighting factor (α) effect

For each specific case of Table 2 (regardless the features, datasets, and face matching problem), the proposed approach SIWEDA+WCCN performs better than its classical SIEDA+WCCN counterpart which obviously demonstrates the effectiveness of the proposed SIWEDA+WCCN method. Moreover, our results demonstrate that SIWEDA+WCCN is able to extract better discriminative features than SIEDA+WCCN.

4.5 Effect of score fusion

Toward to answer the question: What is the best weighting scale of our StatBIF descriptor can be used? We check the complementarity of different weighting scales (i.e., $\alpha = 15$ and $\alpha = 20$) of StatBIF descriptor. The Multi-scale StatBIF (MStatBIF) score fusion is performed using logistic regression method [19]. We experimented with three datasets LFW and YTF for face matching problem and TSKinFace

Table 2 Mean verification accuracy (%) of our SIWEDA+WCCN method using StatBIF descriptor with different weighted factors on Cornell KinFace, UB KinFace, TSKinFace, LFW and YTF datasets

Descriptor	Cornell	UB KinFace			TSKinFace					LFW	YTF
	Mean	Set 1	Set 2	Mean	F-S	F-D	M-S	M-D	Mean	Mean ± std	Mean ± std
MLBP ($\alpha = 1$)	73.83	72.51	69.44	70.98	80.80	79.97	82.26	82.06	81.27	93.97 ± 1.01	70.80 ± 0.58
MLPQ ($\alpha = 1$)	75.25	72.22	70.91	71.57	82.47	80.78	83.72	84.65	82.91	94.30 ± 0.86	71.92 ± 0.52
MBSIF ($\alpha = 1$)	75.95	72.51	70.90	71.71	82.07	81.38	82.65	84.35	82.61	94.43 ± 0.96	70.92 ± 0.51
StatBIF ($l = 3, \alpha = 1$)	76.91	73.22	71.42	72.32	82.37	81.87	82.16	84.75	82.79	92.60 ± 1.12	71.40 ± 0.42
StatBIF ($l = 5, \alpha = 1$)	75.91	73.23	71.66	72.44	82.46	82.87	83.04	86.05	83.61	93.87 ± 0.93	72.40 ± 0.33
StatBIF ($l = 7, \alpha = 1$)	75.28	73.74	71.13	72.44	81.97	81.37	83.82	85.15	83.08	94.00 ± 1.00	71.68 ± 0.44
StatBIF ($l = 9, \alpha = 1$)	75.63	71.74	70.14	70.94	81.97	80.58	83.62	84.25	82.61	94.20 ± 0.92	71.32 ± 0.48
StatBIF ($l = 11, \alpha = 1$)	77.01	72.24	69.65	70.95	82.94	81.37	83.53	83.56	82.85	94.03 ± 0.96	70.96 ± 0.62
StatBIF ($l = 13, \alpha = 1$)	77.71	72.73	70.41	71.57	82.85	79.69	83.43	83.86	82.46	94.27 ± 0.90	70.72 ± 0.62
MLBP ($\alpha = 5$)	78.07	74.48	74.19	74.34	81.49	79.27	82.94	82.36	81.52	94.20 ± 1.00	75.28 ± 0.50
MLPQ ($\alpha = 5$)	77.29	72.46	73.66	73.06	83.64	81.17	84.89	85.95	83.91	95.43 ± 0.80	75.84 ± 0.43
MBSIF ($\alpha = 5$)	78.04	73.44	74.4	73.92	84.61	82.07	85.48	86.35	84.63	95.10 ± 0.88	74.56 ± 0.67
StatBIF ($l = 3, \alpha = 5$)	77.99	73.94	75.39	74.67	83.54	81.97	83.63	86.05	83.80	93.60 ± 0.99	74.48 ± 0.38
StatBIF ($l = 5, \alpha = 5$)	75.91	74.72	74.92	74.82	83.73	82.96	84.59	86.45	84.43	94.73 ± 0.87	77.04 ± 0.61
StatBIF ($l = 7, \alpha = 5$)	75.95	74.91	74.15	74.53	83.82	82.56	85.08	86.15	84.40	95.13 ± 0.86	76.92 ± 0.70
StatBIF ($l = 9, \alpha = 5$)	77.29	75.48	73.44	74.46	83.53	81.17	85.28	85.75	83.93	95.23 ± 0.80	78.56 ± 0.44
StatBIF ($l = 11, \alpha = 5$)	**78.40**	74.94	73.43	74.19	84.31	81.67	84.89	85.36	84.06	94.93 ± 0.81	78.28 ± 0.56
StatBIF ($l = 13, \alpha = 5$)	77.66	74.42	76.89	75.66	83.43	81.47	84.21	84.76	83.47	95.03 ± 0.79	77.52 ± 0.30
MLBP ($\alpha = 10$)	76.65	73.21	75.95	74.58	80.71	77.98	81.96	80.77	80.36	94.30 ± 0.96	75.44 ± 0.44
MLPQ ($\alpha = 10$)	77.23	73.96	74.17	74.07	83.63	80.97	85.08	86.05	83.93	95.73 ± 0.75	76.60 ± 0.41
MBSIF ($\alpha = 10$)	76.91	74.69	76.40	75.55	84.61	81.77	**86.45**	85.95	84.70	95.47 ± 0.80	75.08 ± 0.54
StatBIF ($l = 3, \alpha = 10$)	77.99	74.93	76.64	75.79	83.73	81.86	84.40	85.45	83.86	93.80 ± 0.97	73.68 ± 0.46
StatBIF ($l = 5, \alpha = 10$)	76.63	75.46	76.41	75.94	84.22	82.67	85.08	86.75	84.68	95.20 ± 0.82	77.28 ± 0.71
StatBIF ($l = 7, \alpha = 10$)	75.17	75.65	75.15	75.40	84.21	82.56	85.09	86.05	84.48	95.57 ± 0.78	77.12 ± 0.57
StatBIF ($l = 9, \alpha = 10$)	76.58	75.19	74.66	74.93	83.92	81.67	85.77	85.65	84.25	95.73 ± 0.74	79.24 ± 0.44
StatBIF ($l = 11, \alpha = 10$)	77.64	75.67	75.16	75.42	84.50	82.36	85.97	85.26	84.52	95.33 ± 0.75	78.96 ± 0.54
StatBIF ($l = 13, \alpha = 10$)	76.91	75.14	76.89	76.02	83.14	82.17	84.60	85.36	83.82	95.17 ± 0.75	77.16 ± 0.49
MLBP ($\alpha = 15$)	76.76	72.94	76.19	74.57	80.22	77.18	81.67	80.67	79.94	94.27 ± 0.94	75.40 ± 0.52
MLPQ ($\alpha = 15$)	76.52	74.71	74.67	74.69	83.54	80.97	84.98	85.75	83.81	95.67 ± 0.75	77.48 ± 0.34
MBSIF ($\alpha = 15$)	77.26	75.17	76.39	75.78	84.80	82.27	86.16	86.35	**84.90**	95.57 ± 0.75	75.32 ± 0.57
StatBIF ($l = 3, \alpha = 15$)	77.25	75.69	76.37	76.03	83.93	81.56	84.40	85.05	83.74	93.70 ± 1.00	73.16 ± 0.57
StatBIF ($l = 5, \alpha = 15$)	76.92	75.46	**76.92**	76.19	83.92	**83.16**	84.99	**86.95**	84.76	95.33 ± 0.80	77.16 ± 0.78
StatBIF ($l = 7, \alpha = 15$)	75.17	75.65	75.40	75.53	83.82	82.36	85.28	86.05	84.38	95.90 ± 0.75	77.80 ± 0.55
StatBIF ($l = 9, \alpha = 15$)	76.19	75.17	75.15	75.16	84.02	81.57	85.86	84.96	84.10	**95.97 ± 0.66**	**79.60 ± 0.57**
StatBIF ($l = 11, \alpha = 15$)	76.91	75.16	75.39	75.28	84.41	82.16	86.16	84.16	84.22	95.43 ± 0.80	78.56 ± 0.60
StatBIF ($l = 13, \alpha = 15$)	76.91	75.63	76.16	75.90	83.33	81.57	83.81	85.56	83.57	95.33 ± 0.71	77.32 ± 0.48
MLBP ($\alpha = 20$)	75.88	73.95	76.21	75.08	79.73	76.68	80.89	79.77	79.27	94.17 ± 0.96	75.08 ± 0.62
MLPQ ($\alpha = 20$)	76.16	75.19	74.69	74.94	83.54	81.37	85.76	85.05	83.93	95.43 ± 0.79	77.68 ± 0.35
MBSIF ($\alpha = 20$)	76.56	75.16	76.92	76.04	**85.00**	81.67	85.86	85.95	84.62	95.43 ± 0.79	75.72 ± 0.53
StatBIF ($l = 3, \alpha = 20$)	76.56	75.93	75.87	75.90	84.02	81.56	84.89	85.15	83.91	93.63 ± 1.04	72.20 ± 0.72
StatBIF ($l = 5, \alpha = 20$)	75.15	75.71	76.92	**76.32**	83.83	83.16	85.18	86.55	84.68	95.40 ± 0.81	76.52 ± 0.74
StatBIF ($l = 7, \alpha = 20$)	74.84	**76.16**	75.65	75.91	83.62	82.86	85.67	85.35	84.38	95.80 ± 0.73	77.80 ± 0.44
StatBIF ($l = 9, \alpha = 20$)	76.54	75.68	75.40	75.54	83.73	81.07	85.38	84.86	83.76	95.93 ± 0.66	79.20 ± 0.74
StatBIF ($l = 11, \alpha = 20$)	76.91	74.90	75.14	75.02	84.31	82.07	86.06	83.66	84.03	95.40 ± 0.83	78.12 ± 0.66
StatBIF ($l = 13, \alpha = 20$)	77.25	75.63	76.16	75.90	83.04	81.67	83.62	84.66	83.25	95.33 ± 0.75	77.24 ± 0.46

The numbers in bold represent the best results for each column

Table 3 Mean verification accuracy of MStatBIF with different weighting scales of StatBIF descriptor on LFW, YTF and TSKinFace datasets

Method	LFW	YTF	TSKinFace
StatBIF ($l = 3, \alpha = 15$)	93.70 ± 1.00	73.16 ± 0.57	83.74
StatBIF ($l = 5, \alpha = 15$)	95.33 ± 0.80	77.16 ± 0.78	84.76
StatBIF ($l = 7, \alpha = 15$)	95.90 ± 0.75	77.80 ± 0.55	84.38
StatBIF ($l = 9, \alpha = 15$)	95.97 ± 0.66	79.60 ± 0.57	84.10
StatBIF ($l = 11, \alpha = 15$)	95.43 ± 0.80	78.56 ± 0.60	84.22
StatBIF ($l = 13, \alpha = 15$)	95.33 ± 0.71	77.32 ± 0.48	83.57
StatBIF ($l = 3, \alpha = 20$)	93.63 ± 1.04	72.20 ± 0.72	83.91
StatBIF ($l = 5, \alpha = 20$)	95.40 ± 0.81	76.52 ± 0.74	84.68
StatBIF ($l = 7, \alpha = 20$)	95.80 ± 0.73	77.80 ± 0.44	84.38
StatBIF ($l = 9, \alpha = 20$)	95.93 ± 0.66	79.20 ± 0.74	83.76
StatBIF ($l = 11, \alpha = 20$)	95.40 ± 0.83	78.12 ± 0.66	84.03
StatBIF ($l = 13, \alpha = 20$)	95.33 ± 0.75	77.24 ± 0.46	83.25
MStatBIF ($\alpha = 15$)	96.03 ± 0.66	80.08 ± 0.53	**86.37**
MStatBIF ($\alpha = 20$)	**96.20 ± 0.63**	**80.24 ± 0.47**	86.30

The numbers in bold represent the best results for each column

for kinship verification problem. Table 3 shows the comparison of MStatBIF with different weighting scales of our StatBIF descriptor. As we can see from this table, our MStatBIF shows best and stable performance from the three datasets LFW, YTF and TSKinFace. Furthermore, the Cornell KinFace and UB KinFace datasets are gathered with a mixture of four relations types, Father-Son, Father-Daughter, Mother-Son, and Mother-Daughter pair images, unlike to TSKinFace dataset that subdivided into four subsets (i.e. F-S subset, F-D subset, M-S subset, and M-D subset) which make better scores learning with logistic regression method [26, 27].

4.6 Computational cost

We calculated the computational time needed for the face and kinship verification of one pair face samples using different weighting scales of our SIWEDA+WCCN method. The experiments were implemented using MATLAB 2018a

Table 4 Time Cost (TC), in ms, taken by different weighting factors for the projection of one pair of facial images

Database	Projection and matching				
	$\alpha = 1$	$\alpha = 5$	$\alpha = 10$	$\alpha = 15$	$\alpha = 20$
Cornell KinFace	42.03	39.18	35.64	32.53	31.15
UB KinFace	45.88	41.22	37.45	34.80	30.88
TSKinFace	52.60	48.14	46.39	41.01	39.06
LFW	73.22	71.68	46.21	43.56	39.25
YTF	67.95	65.41	63.55	59.96	57.26

Table 5 Comparison verification accuracy of StatBIF+SIWEDA +WCCN with image restricted setting (no outside training data was used) on LFW dataset

Method	Mean accuracy $\pm$ standard error (%)
Eigenfaces, original [55]	60.02 ± 0.79
Nowak2, original [40]	72.45 ± 0.40
Nowak2, funneled [22]	73.93 ± 0.49
Hybrid descriptor-based, funneled [58]	78.47 ± 0.51
3x3 Multi-Region Histograms [51]	72.95 ± 0.55
Pixels/MKL, funneled [47]	68.22 ± 0.41
V1-like/MKL, funneled [47]	79.35 ± 0.55
APEM (fusion), funneled [29]	84.08 ± 1.20
MRF-MLBP [1]	79.08 ± 0.14
Fisher vector faces [52]	87.47 ± 1.49
Eigen-PEP 30]	88.97 ± 1.32
MRF-MBSIF-CSKDA [2]	93.63 ± 1.27
MRF-Fusion-CSKDA [2]	95.89 ± 1.94
POP-PEP [28]	91.10 ± 1.47
DDML [20]	90.68 ± 1.41
LM^3L[21]	89.57 ± 1.53
Discriminative deep multi-metric learning [33]	93.28 ± 0.39
CA-LBFL [14]	92.75 ± 1.13
StatBIF-SIWEDA-WCCN (Our)	**96.20 ± 0.63**

The numbers in bold represent the best results

on a PC with an Intel Core i7 2.00 GHz CPU and 8 GB of RAM. The feature extraction for the MLBP, MLPQ, MBSIF and StatBIF descriptors takes 7.3 ms, 37.6 ms, 27.3 ms and 23.1 ms for each sample, respectively. Furthermore, in addition to its robustness, our StatBIF descriptor got the second rank in term of time cost compared with the three well known descriptors. In training step (offline), the estimation of the projection matrices is performed only once. In the online step, we evaluate the time cost needed by each method to project and match the test pair which is provided in Table 4 in ms. This table shows that the best performing method, SIWEDA, runs faster compared with the SIEDA method (i.e. α=1). Furthermore, we see that the time cost of feature extraction is negligible compared to the projection and matching time.

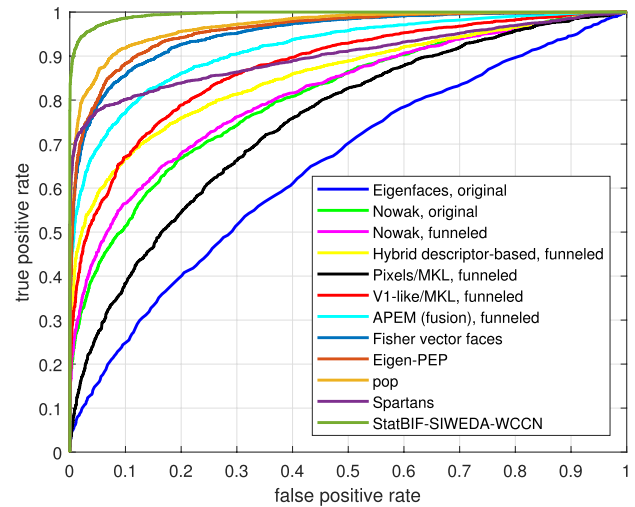## 4.7 Comparison with the results of the state of the art

- *Matching face pairs in the Wild* Table 5 shows the comparison of our results under restricted setting protocol with the state of the art methods on the LFW dataset. The corresponding ROC curves of our framework and the state of the art methods on LFW dataset are depicted in Fig. 4. The best achieved verification accuracy of our framework is 96.20%,

Table 6 Comparison verification accuracy of StatBIF+SIWEDA +WCCN with image restricted setting (no outside training data was used) on YTF dataset

Method	Mean accuracy ± standard error (%)
MBGS L2 mean [57]	76.40 ± 1.80
APEM-FUSION [29]	79.10 ± 1.50
STFRD+PMML [10]	79.50 ± 2.50
VSOF+OSS [39]	79.70 ± 1.80
DDML [20]	82.34 ± 1.47
LM^{3}L [21]	81.30 ± 1.20
Discriminative deep multi-metric learning [33]	82.54 ± 1.58
CA-LBFL [14]	**83.30 ± 1.30**
StatBIF-SIWEDA-WCCN (Our)	**80.24 ± 0.47**

The numbers in bold represent the best results and our approach results

Fig. 4 ROC curve of StatBIF-SIWEDA-WCCN and other state-of-the-art methods on the LFW dataset under image restricted setting
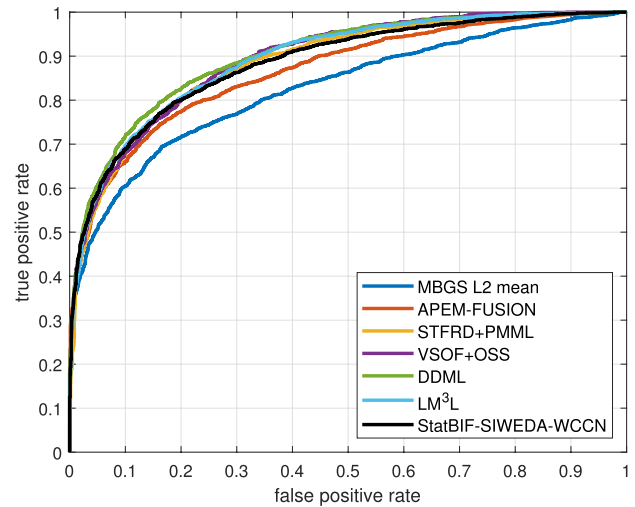
Fig. 5 ROC curve of StatBIF-SIWEDA-WCCN and other state-of-the-art methods on the YTF dataset under image restricted setting

which is very close to the human performance (i.e., 97.53% on LFW dataset). Furthermore, 80 positive and negative pairs facial images from the 6000 pairs (almost average of 8 positive and negative pairs for each fold) that are misclassified by our approach compared to human performance. Comparing to the state of the art, our result is the first. The framework MRF-Fusion-CSKDA [2] achieves currently the second rank on the LFW dataset in term of verification accuracy. MRF-Fusion-CSKDA is resulted by the fusion of three descriptors which extracted at multi-scale features, MSLBP, MSBSIF, and MSLPQ, and kernel methods. Differently, we utilized only one descriptor (StatBIF), with an effective approach, SIWEDA+WCCN, achieving first rank. Thus, our framework is less complicated and more efficiently computational than the second ranking method. Furthermore, Table 6 shows the comparison of our results under restricted setting protocol with the state of the art methods on the YTF dataset. The corresponding ROC curves of our framework and the state of the art methods on YTF dataset are depicted in

Table 7 Comparison verification accuracy of StatBIF+SIWEDA+WCCN with state of the art on the Cornell KinFace, UB KinFace and TSKinFace datasets

Method	Cornell	UB KinFace	TSKinFace
Pictorial structure model [16]	70.67	/	/
Transfer subspace learning [61]	/	68.50	/
Neighborhood repulsed metric learning [34]	71.60	67.05	/
Discriminative multimetric learning [62]	73.50	72.25	/
Prototype discriminative feature learning [63]	71.90	67.30	/
Relative symmetric bilinear model [49]	/	/	81.85
BSIF-HSV [60]	/	/	81.19
Discriminative deep multi-metric learning [33]	/	/	84.15
MHDL3 - {HOG + Color + LPQ} [35]	76.60	/	/
Heterogeneous similarity learning [48]	68.40	56.20	/
StatBIF-SIWEDA-WCCN (Our)	**78.40**	**76.32**	**86.37**

The numbers in bold represent the best results for each column

Fig. 5. Thus, our framework got a competitive performance compared to state of the art methods on YTF dataset under restricted setting protocol.

- Kinship Verification in the wild Table 7 compares the proposed approach with the state of the art methods on the Cornell KinFace, UB KinFace and TSKinFace datasets. We notice that the best verification accuracy of our framework achieves 78.40% on Cornell KinFace, 76.32% on UB KinFace and 86.37% on TSKinFace. As can be seen from these results, our approach outperforms the other state of the art methods on two kinship datasets. In addition, we can see that SIWEDA+WCCN improves with a significant margin the kinship performance (more than 4% improvement on Cornell KinFace and UB KinFace). From TSKinFace dataset, we see that the proposed SIWEDA+WCCN method obtains the first best results for all relationships.

Our approach vs. DeepFace [54] (**using 4.4M outside data)** We compared our approach using the provided data only on LFW dataset with DeepFace [54] method which used 4.4 Millions outside facial images belonging to more than 4000 identities for training the network model. We got a competitive performance compared with DeepFace method under restricted setting on LFW dataset (we refer that our StatBIF+SIWEDA+WCCN method and Deep-Face method got a verification accuracies of 96.20% and 97.15%, respectively). Moreover, the authors in [54] combine three networks feeding by three types of inputs (i.e. a 3D-aligned three-channels (RGB) facial images; a gray facial images plus image gradient magnitude and orientation; and a 2D-aligned RGB facial images) of size 152 by 152, unlike to our approach (StatBIF+SIWEDA+WCCN) which used gray input facial images of size 64 by 64 with no preprocessing stage applied to the facial images (a raw facial images was used).

Our approach vs. Deep multi-metric learning [33] In this comparison, we focus on three datasets LFW, YTF and TSKinFace. We see that our approach improves the performance with about 3% from LFW dataset, got a competitive performance from YTF dataset, and improves the performance with about 2% from TSKinFace dataset compared with DDMML. The Discriminative Deep Multi-Metric Learning (DDMML) [33] method used the combination of multiple features (multi-view features) to describe the facial images. The work of Lu et al. [33] adopts the different sizes to extract the features of the input facial images for each dataset (i.e. 80×150 for LFW dataset, 100×100 for YTF dataset, and 64×64 for TSKinFace dataset). Furthermore, they extract different features for each dataset, which are six original and square root features of Sparse SIFT (SSIFT) [17] on the "funneled" LFW dataset, histogram of oriented gradients (HOG) [11] on the LFW-a dataset and high-dimensional LBP (HDLBP) [9] on the original LFW (i.e. they used the combination of six of original and square root features

applied on three versions of LFW dataset). For YTF dataset, three features description including LBP [41], Center-Symmetric LBP (CSLBP) 57] and Four-Patch LBP (FPLBP) [58] are used as in [57]. For TSKinFace dataset, they used four descriptors LBP, Dense SIFT (DSIFT), HOG and LPQ. On the other hand, our approach used efficient methods (SIWEDA+WCCN) using the proposed descriptor (StatBIF) only. Furthermore, our approach used the same parameter settings of the all four used descriptors in our work for both face matching and kinship verification problems (i.e. for each dataset the input size is 64×64 and we used the same features extraction parameters for MLBP, MBSIF, MLPQ, and MStatBIF descriptors on each dataset) attains stable, robust and good performances. Therefore, the DDMML method accepts only a specific input size of facial images and adapted descriptors (under human surveillance/guidance learning) for each dataset, unlike SIWEDA+WCCN do. Besides, this means that the DDMML method has weak learning compared with our SIWEDA+WCCN method for face and kinship verification in all scenarios.

5 Conclusion

In this work, we proposed an approach to the problem of face and kinship verification through the weighting factor (α) of SIWEDA+WCCN method in various texture descriptors (MSLBP, MSLPQ, MSBSIF and the proposed StatBIF), we experimented with four weighting factors ($\alpha = 5, \alpha = 10, \alpha = 15$ and $\alpha = 20$) in addition to the classical value ($\alpha = 1$). Thorough experiments are performed on five datasets in the wild, namely the LFW, the YTF, the Cornell KinFace, the UB KinFace and the TSKinFace. The obtained results show the effectiveness of using our Stat-BIF descriptor over the different values of weighting factors superior to one ($\alpha > 1$) for face and kinship verification compared with the classical value of weighting factor ($\alpha = 1$). These results point out the importance of StatBIF-SIWEDA-WCCN approach for face and kinship verification in all scenarios. Additionally, our results compare favorably against the recent approaches in the literature on the benchmark datasets. Moreover, our approach demonstrates better results than the discriminative deep multi-metric learning method on LFW and TSKinFace datasets.

References

1. Arashloo SR, Kittler J (2013) Efficient processing of mrfs for unconstrained-pose face recognition. In: 2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS), pp 1–8. https://doi.org/10.1109/BTAS.2013.6712721
2. Arashloo SR, Kittler J (2014) Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised

statistical image features. IEEE Trans Inf Forensics Secur 9(12):2100–2109. https://doi.org/10.1109/TIFS.2014.2359587

3. Ayesha S, Hanif MK, Talib R (2020) Overview and comparative study of dimensionality reduction techniques for high dimensional data. Inf Fusion 59:44–58. https://doi.org/10.1016/j.inffus.2020.01.005

4. Barkan O, Weill J, Wolf L, Aronowitz H (2013) Fast high dimensional vector multiplication face recognition. In: 2013 IEEE international conference on computer vision, pp 1960–1967. https://doi.org/10.1109/ICCV.2013.246

5. Bekhouche S, Ouafi A, Dornaika F, Taleb-Ahmed A, Hadid A (2017) Pyramid multi-level features for facial demographic estimation. Expert Syst Appl 80:297–310. https://doi.org/10.1016/j.eswa.2017.03.030

6. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720. https://doi.org/10.1109/34.598228

7. Best-Rowden L, Bisht S, Klontz JC, Jain AK (2014) Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. In: IEEE International Joint Conference on Biometrics, pp. 1–8 (2014). https://doi.org/10.1109/BTAS.2014.6996296

8. Chakrabarti A, Rajagopalan AN, Chellappa R (2007) Super-resolution of face images using kernel pca-based prior. IEEE Trans Multimed 9(4):888–892. https://doi.org/10.1109/TMM.2007.893346

9. Chen D, Cao X, Wen F, Sun J (2013) Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp 3025–3032. https://doi.org/10.1109/CVPR.2013.389

10. Cui Z, Li W, Xu D, Shan S, Chen X (2013) Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp 3554–3561. https://doi.org/10.1109/CVPR.2013.456

11. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1, pp 886–893. https://doi.org/10.1109/CVPR.2005.177

12. Dehak N, Kenny PJ, Dehak R, Dumouchel P, Ouellet P (2011) Front-end factor analysis for speaker verification. IEEE Trans Audio Speech Lang Process 19(4):788–798. https://doi.org/10.1109/TASL.2010.2064307

13. Duan Y, Lu J, Feng J, Zhou J (2017) Learning rotation-invariant local binary descriptor. IEEE Trans Image Process 26(8):3636–3651. https://doi.org/10.1109/TIP.2017.2704661

14. Duan Y, Lu J, Feng J, Zhou J (2018) Context-aware local binary feature learning for face recognition. IEEE Trans Pattern Anal Mach Intell 40(5):1139–1153. https://doi.org/10.1109/TPAMI.2017.2710183

15. Duan Y, Lu J, Wang Z, Feng J, Zhou J (2017) Learning deep binary descriptor with multi-quantization. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 4857–4866. https://doi.org/10.1109/CVPR.2017.516

16. Fang R, Tang KD, Snavely N, Chen T (2010) Towards computational models of kinship verification. In: 2010 IEEE international conference on image processing, pp 1577–1580. https://doi.org/10.1109/ICIP.2010.5652590

17. Guillaumin M, Verbeek J, Schmid C (2009) Is that you? metric learning approaches for face identification. In: 2009 IEEE 12th international conference on computer vision, pp 498–505. https://doi.org/10.1109/ICCV.2009.5459197

18. Haghighat M, Abdel-Mottaleb M, Alhalabi W (2016) Fully automatic face normalization and single sample face recognition in unconstrained environments. Expert Syst Appl 47:23–34. https://doi.org/10.1016/j.eswa.2015.10.047

19. Harrell FE Jr (2015) Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer, Berlin

20. Hu J, Lu J, Tan Y (2014) Discriminative deep metric learning for face verification in the wild. In: 2014 IEEE conference on computer vision and pattern recognition, pp 1875–1882. https://doi.org/10.1109/CVPR.2014.242

21. Hu J, Lu J, Yuan J, Tan YP (2014) Large margin multi-metric learning for face and kinship verification in the wild. IEEE Trans Circ Syst Video Technol. https://doi.org/10.1007/978-3-319-16811-1_17

22. Huang GB, Jain V, Learned-Miller E (2007) Unsupervised joint alignment of complex images. In: 2007 IEEE 11th international conference on computer vision, pp 1–8. https://doi.org/10.1109/ICCV.2007.4408858

23. Huang GB, Lee H, Learned-Miller E (2012) Learning hierarchical representations for face verification with convolutional deep belief networks. In: 2012 IEEE conference on computer vision and pattern recognition, pp 2518–2525. https://doi.org/10.1109/CVPR.2012.6247968

24. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech. Rep. 07–49. University of Massachusetts, Amherst

25. Kannala J, Rahtu E (2012) Bsif: binarized statistical image features. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), pp 1363–1366

26. Laiadi O, Ouamane A, Benakcha A, Taleb-Ahmed A, Hadid A (2019) Learning multi-view deep and shallow features through new discriminative subspace for bi-subject and tri-subject kinship verification. Appl Intell 49(11):3894–3908

27. Laiadi O, Ouamane A, Boutellaa E, Benakcha A, Taleb-Ahmed A, Hadid A (2019) Kinship verification from face images in discriminative subspaces of color components. Multimed Tools Appl 78(12):16465–16487

28. Li H, Hua G (2015) Hierarchical-pep model for real-world face recognition. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 4055–4064. https://doi.org/10.1109/CVPR.2015.7299032

29. Li H, Hua G, Lin Z, Brandt J, Yang J (2013) Probabilistic elastic matching for pose variant face verification. In: 2013 IEEE conference on computer vision and pattern recognition, pp 3499–3506. https://doi.org/10.1109/CVPR.2013.449

30. Li H, Hua G, Shen X, Lin Z, Brandt J (2015) Eigen-PEP for video face recognition. Springer, Cham, pp 17–33. https://doi.org/10.1007/978-3-319-16811-1_2

31. Liu C (2014) Discriminant analysis and similarity measure. Pattern Recognit 47(1):359–367. https://doi.org/10.1016/j.patcog.2013.06.023

32. Lu GF, Wang Y, Zou J, Wang Z (2018) Matrix exponential based discriminant locality preserving projections for feature extraction. Neural Netw 97:127–136. https://doi.org/10.1016/j.neunet.2017.09.014

33. Lu J, Hu J, Tan YP (2017) Discriminative deep metric learning for face and kinship verification. IEEE Trans Image Process 26(9):4269–4282. https://doi.org/10.1109/TIP.2017.2717505

34. Lu J, Zhou X, Tan YP, Shang Y, Zhou J (2014) Neighborhood repulsed metric learning for kinship verification. IEEE Trans Pattern Anal Mach Intell 36(2):331–345. https://doi.org/10.1109/TPAMI.2013.134

35. Mahpod S, Keller Y (2018) Kinship verification using multiview hybrid distance learning. Comput Vis Image Underst 167:28–36. https://doi.org/10.1016/j.cviu.2017.12.003

36. Mao Q, Rao Q, Yu Y, Dong M (2017) Hierarchical bayesian theme models for multipose facial expression recognition. IEEE Trans Multimed 19(4):861–873. https://doi.org/10.1109/TMM.2016.2629282

37. Marsico MD, Nappi M, Riccio D, Wechsler H (2013) Robust face recognition for uncontrolled pose and illumination changes. IEEE Trans Syst Man Cybern Syst 43(1):149–163. https://doi.org/10.1109/TSMCA.2012.2192427

38. Meina K, Shan S, Chen X (2011) Side-information based linear discriminant analysis for face recognition. In: Proc. BMVC, pp 125.1–125.0. https://doi.org/10.5244/C.25.125

39. Méndez-Váizquez, H., Martínez-Díaz, Y., Chai, Z.: Volume structured ordinal features with background similarity measure for video face recognition. In: 2013 International conference on biometrics (ICB), pp 1–6 (2013). https://doi.org/10.1109/ICB.2013.6612990

40. Nowak E, Jurie F (2007) Learning visual similarity measures for comparing never seen objects. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–8. https://doi.org/10.1109/CVPR.2007.382969

41. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution grayscale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987. https://doi.org/10.1109/TPAMI.2002.1017623

42. Ojansivu V, Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. Springer, Berlin, pp 236–243. https://doi.org/10.1007/978-3-540-69905-7_27

43. Ouamane A, Bengherabi M, Hadid A, Cheriet M (2015) Side-information based exponential discriminant analysis for face verification in the wild. In: 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), vol 02, pp 1–6. https://doi.org/10.1109/FG.2015.7284837

44. Ouamane A, Chouchane A, Boutellaa E, Belahcene M, Bourennane S, Hadid A (2017) Efficient tensor-based 2d+3d face verification. IEEE Trans Inf Forensics Secur 12(11):2751–2762. https://doi.org/10.1109/TIFS.2017.2718490

45. Ouamane A, Messaoud B, Guessoum A, Hadid A, Cheriet M (2014) Multi scale multi descriptor local binary features and exponential discriminant analysis for robust face authentication. In: 2014 IEEE international conference on Iimage processing (ICIP), pp 313–317. https://doi.org/10.1109/ICIP.2014.7025062

46. Pang Y, Wang S, Yuan Y (2014) Learning regularized lda by clustering. IEEE Trans Neural Netw Learn Syst 25(12):2191–2201. https://doi.org/10.1109/TNNLS.2014.2306844

47. Pinto N, DiCarlo JJ, Cox DD (2009) How far can you get with a modern face recognition test set using only simple features? In: 2009 IEEE conference on computer vision and pattern recognition, pp 2591–2598. https://doi.org/10.1109/CVPR.2009.5206605

48. Qin X, Liu D, Wang D (2018) Heterogeneous similarity learning for more practical kinship verification. Neural Process Lett 47(3):1253–1269. https://doi.org/10.1007/s11063-017-9694-3

49. Qin X, Tan X, Chen S (2015) Tri-subject kinship verification: Understanding the core of a family. IEEE Trans Multimed 17(10):1855–1867. https://doi.org/10.1109/TMM.2015.2461462

50. Raudys SJ, Jain AK (1991) Small sample size effects in statistical pattern recognition: recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell 13(3):252–264. https://doi.org/10.1109/34.75512

51. Sanderson C, Lovell BC (2009) Multi-region probabilistic histograms for robust and scalable identity Inference. Springer, Berlin, pp 199–208. https://doi.org/10.1007/978-3-642-01793-3_21

52. Simonyan K, Parkhi OM, Vedaldi A, Zisserman A (2013) Fisher vector faces in the wild. BMVC 2(3):4

53. Swets DL, Weng JJ (1996) Using discriminant eigenfeatures for image retrieval. IEEE Trans Pattern Anal Mach Intell 18(8):831–836. https://doi.org/10.1109/34.531802

54. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: 2014 IEEE conference on computer vision and pattern recognition, pp 1701–1708. https://doi.org/10.1109/CVPR.2014.220

55. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: Proceedings 1991 IEEE computer society conference on computer vision and pattern recognition, pp 586–591. https://doi.org/10.1109/CVPR.1991.139758

56. Wei W, Dai H, Liang W (2020) Exponential sparsity preserving projection with applications to image recognition. Pattern Recognit. https://doi.org/10.1016/j.patcog.2020.107357

57. Wolf L, Hassner T, Maoz I (2011) Face recognition in unconstrained videos with matched background similarity. CVPR 2011:529–534. https://doi.org/10.1109/CVPR.2011.5995566

58. Wolf L, Hassner T, Taigman Y (2008) Descriptor based methods in the wild. In: Real-life images workshop at the European conference on computer vision (ECCV). http://www.openu.ac.il/home/hassner/projects/Patchlbp

59. Wu G, Feng T, Zhang L, Yang M (2017) Inexact implementation using krylov subspace methods for large scale exponential discriminant analysis with applications to high dimensionality reduction problems. Pattern Recognit 66:328–341. https://doi.org/10.1016/j.patcog.2016.08.020

60. Wu X, Boutellaa E, López MB, Feng X, Hadid A (2016) On the usefulness of color for kinship verification from face images. In: 2016 IEEE International workshop on information forensics and security (WIFS), pp 1–6. https://doi.org/10.1109/WIFS.2016.7823901

61. Xia S, Shao M, Luo J, Fu Y (2012) Understanding kin relationships in a photo. IEEE Trans Multimed 14(4):1046–1056. https://doi.org/10.1109/TMM.2012.2187436

62. Yan H, Lu J, Deng W, Zhou X (2014) Discriminative multimetric learning for kinship verification. IEEE Trans Inf Forensics Secur 9(7):1169–1178. https://doi.org/10.1109/TIFS.2014.2327757

63. Yan H, Lu J, Zhou X (2015) Prototype-based discriminative feature learning for kinship verification. IEEE Trans Cybern 45(11):2535–2545. https://doi.org/10.1109/TCYB.2014.2376934

64. Yu W, Zhao C (2018) Sparse exponential discriminant analysis and its application to fault diagnosis. IEEE Trans Ind Electron 65(7):5931–5940. https://doi.org/10.1109/TIE.2017.2782232

65. Yuan S, Mao X (2018) Exponential elastic preserving projections for facial expression recognition. Neurocomputing 275:711–724. https://doi.org/10.1016/j.neucom.2017.08.067

66. Zhang T, Fang B, Tang YY, Shang Z, Xu B (2010) Generalized discriminant analysis: a matrix exponential approach. IEEE Trans Syst Man Cybern Part B 40(1):186–197. https://doi.org/10.1109/TSMCB.2009.2024759