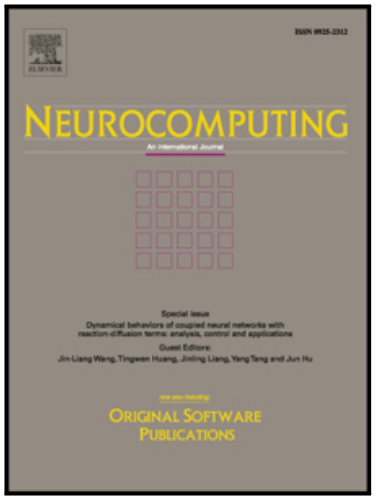
Communicated by Dr Xiaoming Liu

Accepted Manuscript



Multilinear Side-Information based Discriminant Analysis for Face and Kinship Verification in the Wild

Mohcene Bessaoudi, Abdelmalik Ouamane, Mebarka Belahcene, Ammar Chouchane, Elhocine Boutellaa, Salah Bourennane

PII: S0925-2312(18)31124-X
DOI: https://doi.org/10.1016/j.neucom.2018.09.051
Reference: NEUCOM 19984

To appear in: *Neurocomputing*

Received date: 6 July 2017
Revised date: 31 May 2018
Accepted date: 24 September 2018

Please cite this article as: Mohcene Bessaoudi, Abdelmalik Ouamane, Mebarka Belahcene, Ammar Chouchane, Elhocine Boutellaa, Salah Bourennane, Multilinear Side-Information based Discriminant Analysis for Face and Kinship Verification in the Wild, *Neurocomputing* (2018), doi: https://doi.org/10.1016/j.neucom.2018.09.051

Highlights

- We propose a novel method, Multilinear Side-Information based Discriminant Analysis (MSIDA), for dimensionality reduction and classification of tensor data, when the data full class label is missing. MSIDA projects the input face tensor into a new multilinear subspace in which the margin between samples belonging to different classes is enlarged while the margin within samples belonging to same classes is reduced. Additionally, MSIDA reduces the dimension of each tensor mode.

- On the face description level, we propose a new representation based on high order tensors. This representation combines different local descriptors, extracted at different scales, providing better discrimination. The proposed tensor representation is regarded as a new way for fusing local descriptors.

- We empirically evaluate the proposed approach for face based identity and kinship verification on four challenging face benchmarks (LFW, Cornell KinFace, UB KinFace and TSkinface). Comparison against the state-of the-art methods demonstrates the efficiency of our approach.

Multilinear Side-Information based Discriminant Analysis for Face and Kinship Verification in the Wild

Mohcene Bessaoudi[a], Abdelmalik Ouamane[a], Mebarka Belahcene[a], Ammar Chouchane[a], Elhocine Boutellaa[b], Salah Bourennane[c]

[a] *University of Biskra, Algeria.*
[b] *Center for Machine Vision and Signal Analysis, University of Oulu, Finland.*
[c] *Institut Fresnel, Université de Marseille, France.*

Abstract

This paper presents a new approach for face and kinship verification under unconstrained environments. The proposed approach is based on high order tensor representation of face images. The face tensor is built based on local descriptors extracted at multiscales. Besides, we formulate a novel Multilinear Side-Information based Discriminant Analysis (MSIDA) to handle the weakly supervised multilinear subspace projection and classification. Using only the weak label information, MSIDA projects the input face tensor in a new subspace in which the discrimination is improved and the dimension of each tensor mode is reduced simultaneously. Experimental evaluation on four challenging face databases (LFW, Cornell KinFace, UB KinFace and TSKinface) demonstrates that the proposed approach significantly outperforms the current state of the art.

Keywords: Tensor analysis, multilinear subspace projection, multilinear side-information based discriminant analysis, face verification, kinship verification.

1. Introduction

In the last years, automatic face recognition under controlled environments has attained satisfying performances even in large scale biometric systems. However, the performance of these systems is still weak in challenging and unconstrained settings, referred to as "in the wild", characterized by, for e.g., low image quality, blurred image, strong environment illuminations and head poses changes, partial occlusions, etc. In this work, we focus on

the problem of matching pairs of face images [1, 2, 3] in unconstrained environments [4]. Particularly, the targeted applications consists of checking i) whether a pair of facial images are of the same person or not (identity verification) and ii) whether a pair of facial images are of two persons of the same family or not (kinship verification) [5, 6, 7]. In such applications, the available data for training is weakly labeled. The only provided information is whether the pair of face images is a match (positive pair) or a non-match (negative pair), while no access to the identity of the persons is provided.

Among the widely applied approaches for face recognition are the subspace transformation techniques. The objective of these algorithms is to transform the feature space into a lower, and often more discriminative, subspace leading to better classification accuracies [8, 9]. The pioneer approaches of this category are the two popular linear dimensionality reduction methods, Principal Component Analysis (PCA) [10] and linear discriminant analysis (LDA) [11]. Such algorithms usually vectorize the face images yielding a high-dimensional feature vector [12]. The image vectorization also causes losing the valuable face structural information embedded in the pixels' positions. To solve this issue, high order representations of data [13, 14, 15] in multilinear subspaces are applied to the face recognition problem achieving better performances. The multilinear approaches also allow for a multifactor representation of face image samples, where each factor corresponds to an entry of the multiarray. An illustrative example of face representation by tensors of different orders is depicted in Fig. 1. In which, Fig. 1 (a) the face image features are represented as a vector, Fig. 1 (b) represents the face image features as a 2^{nd}-order tensor, and Fig. 1 (c) is a 3^{rd}-order tensor of a face images.

In this work, we tackle the problem of face pair matching based on tensor representation in the context of weakly labeled training data. The main contributions of this paper are: i) We propose a novel method, Multilinear Side-Information based Discriminant Analysis (MSIDA), for dimensionality reduction and classification of tensor data, when the data full class label is missing. MSIDA projects the input face tensor into a new multilinear subspace in which the margin between samples belonging to different classes is enlarged while the margin within samples belonging to same classes is reduced. Additionally, MSIDA reduces the dimension of each tensor mode. ii) On the face description level, we propose a new representation based on high order tensors. This representation combines different local descriptors, extracted at different scales, providing better discrimination. The proposed

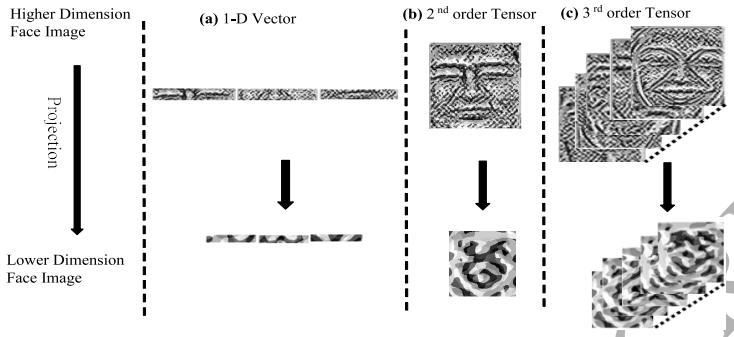


Figure 1: Illustration of face representation and dimensionality reduction with different tensor orders.

tensor representation is regarded as a new way for fusing local descriptors. iii) We empirically evaluate the proposed approach for face based identity and kinship verification on four challenging face benchmarks (LFW, Cornell KinFace, UB KinFace and TSkinface). Comparison against the state-of the-art methods demonstrates the efficiency of our approach.

The rest of our paper is organized as follows. In Section 2, we discuss the related works. In Section 3, we describe the proposed Multilinear Side-Information based Discriminant Analysis. In Section 4, the proposed face matching approach is presented. Section 5 details the experimental evaluation, results and analysis. The conclusion and the future work are given in Section 6.

2. Related work

To manipulate tensors, the linear dimensionality reduction approaches, such as PCA and LDA, have been extended to multilinear subspace algorithms that operate directly on the multidimensional representation instead of their vectorized forms [12, 16, 13, 14]. These approaches define a multilinear transformation that maps the original tensor into a lower dimensional tensor while simultaneously enhancing the disparity among the samples and conserving their intrinsic structural information [14]. For instance, the multilinear principal component analysis (MPCA) [13] is a multilinear extension of the PCA. On the other side, LDA has been extended to a number of multilinear variants. In Multilinear Discriminant Analysis (MDA) [15] multiple interrelated subspaces collaborate to discriminate between different classes.

Different algorithms have been proposed in the literature to reduce the dimensionality of the tensor data. Liu et al. [10] proposed tensor rank-one decomposition and graph preserving criterion dimensionality reduction algorithm for multi-dimensional facial expressions recognition. A marginal neighboring graph is defined to describe the pairwise inter-class boundaries, and a differential formed objective function is adopted to ensure convergence. By seeking the rank-one tensors, the algorithm enhances the pairwise inter-class margins and meanwhile preserves the intra-class local manifold structure. A sparse version of the multilinear discriminant analysis is derived by Lai *et al.* [12]. They introduced the L_1 and L_2 norms into the objective function to obtain multiple interrelated sparse discriminant subspaces for feature extraction. The k-mode optimization technique and the L_1 norm sparse regression are combined to iteratively learn the optimal sparse discriminant subspace along different modes of the tensors. Lu *et al.* [14] proposed uncorrelated multilinear discriminant analysis (UMLDA) for the recognition of tensor objects. Because they contain minimum redundancy and they are independent, the uncorrelated features are useful for recognition tasks. The method extracts the uncorrelated discriminative features directly from tensorial data by solving a tensor-to-vector projection. The solution consists of sequential iterative processes by alternating the projections. Furthermore, an adaptive regularization procedure is incorporated to enhance the performance in the small sample size scenario.

In order to operate, the supervised dimensionality reduction approaches require the full class label information of each data sample. However, this condition cannot always be satisfied, especially under unconstrained conditions where data is acquired from Internet. In such cases, we usually only have access to limited information, such as whether a pair of images is of the same class. An alternative solution to this problem is given by Kan *et al.* [17]. The authors proposed a discriminative dimensionality reduction method named Side-Information based Linear Discriminant Analysis (SILD). SILD directly calculates the within-class and between-class scatter matrices using the side-information (image pair labels). Ouamane *et al.* [18] proposed Exponential Discriminant Analysis (SIEDA) by combining Side-Information Linear Discriminant and Exponential Discriminant Analysis (EDA).

In this work, we propose a multilinear extension of SILD for tensor data subspace analysis. We call the new algorithm Multilinear Side-Information based Discriminant Analysis (MSIDA) and use it for solving the problem of matching pairs of face images, characterized by weakly labeled data. In our

framework, the set of facial images are represented as a 3rd-order tensor built using histograms of two local face descriptors, Multiscale Local Phase Quantization (MSLPQ) [19] and Multiscale Binarized Statistical Image Features (MSBSIF) [20].

3. Multilinear Side-Information based Discriminant Analysis

This section introduces the proposed Multilinear Side-Information based Discriminant Analysis (MSIDA). Firstly, we present the notations and some basics of tensor algebra. Then, we briefly recall the original Side-Information based Linear Discriminant Analysis (SILD). Finally, the proposed extension MSIDA is presented with detailed mathematical formulation.

3.1. Notations and basics of tensor algebra

This part recalls the concepts relevant to the multilinear algebra. For the clarity of mathematical formulations, the following notations of variables are considered throughout the paper. Lowercase and uppercase letters, e.g., i, j, I, J, denote scalars; bold lowercase letters, e.g. $\mathbf{x}$, $\mathbf{y}$, $\mathbf{u}$, $\mathbf{v}$, denote vectors; italic uppercase letters, e.g. U, S, G, H, denote matrices; and bold italic uppercase letters, e.g. $\boldsymbol{X}$, $\boldsymbol{Y}$, denote the tensors.

A tensor is a multidimensional array [12, 15] with each entry of the array is referred to mode and the number of entries is the tensor order. Thus, an N$^{\text{th}}$-order tensor is denoted as $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{N}}$, where I_K, $1 \leq \mathrm{K} \leq \mathrm{N}$, indicates the dimension of the K$^{\text{th}}$ mode of the tensor. The elements of the tensor $\boldsymbol{X}$ are denoted as $\mathrm{x}_{\mathrm{i}_1 \mathrm{i}_2 \cdots \mathrm{i}_\mathrm{N}}$, where $1 \leq \mathrm{i}_\mathrm{K} \leq \mathrm{I}_\mathrm{K}$.

The following definitions introduce some mathematical tools used to manipulate the high order tensors.

Definition 1. the inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle$ of two tensors $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{N}}$ which have the same order and dimensions is defined by:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{\mathrm{i}_1=1,\ldots,\mathrm{i}_\mathrm{N}=1}^{\mathrm{I}_1,\ldots,\mathrm{I}_\mathrm{N}} \mathrm{X}_{\mathrm{i}_1,\ldots,\mathrm{i}_\mathrm{N}} \mathrm{Y}_{\mathrm{i}_1,\ldots,\mathrm{i}_\mathrm{N}} \tag{1}$$

Hence, the norm of a tensor $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{N}}$ is defined as $\|\boldsymbol{X}\|_\mathrm{F} = \sqrt{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}$, and the Euclidean distance between two tensors $\boldsymbol{X}$, $\boldsymbol{Y} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{N}}$ is defined as $D(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_\mathrm{F}$.

Definition 2. the k-mode flattening of a tensor $\boldsymbol{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ into an unfolding matrix $X^{(k)} \in \mathbb{R}^{I_k \times \prod_{i \neq k} I_i}$ is defined by:

$$X^{(k)} \Leftarrow_k \boldsymbol{X} \tag{2}$$

where

$$X^{(k)}_{i_k j} = \boldsymbol{X}_{i_1,\ldots,i_N}, j = 1 + \sum_{I=1, I \neq k}^{N} (i_1 - 1) \prod_{o=I+1, o \neq k}^{N} I_o \tag{3}$$

Definition 3. the k-mode product of a tensor $\boldsymbol{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and a matrix $G \in \mathbb{R}^{I'_k \times I_k}$(k=1,2,...,N) is an $I_1 \times I_2 \times \ldots I_{k-1} \times I'_k \times I_{k+1} \times \ldots \times I_N$ tensor denoted by:

$$\boldsymbol{Y} = \boldsymbol{X} \times_k G \tag{4}$$

Fig. 2 illustrates an example of 1-mode vector product of third-order tensor $\boldsymbol{X} \in \mathbb{R}^{12 \times 7 \times 4}$ with matrix $G^T \in \mathbb{R}^{4 \times 12}$, where the result is a tensor $\boldsymbol{X} \times_1 G^T \in \mathbb{R}^{4 \times 7 \times 4}$

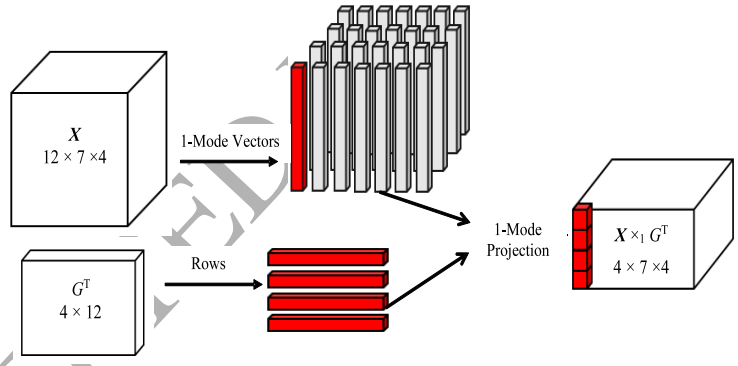


Figure 2: Visual illustration of 1-mode vector product of third-order tensor $\boldsymbol{X} \in \mathbb{R}^{12 \times 7 \times 4}$ with matrix $G^T \in \mathbb{R}^{4 \times 12}$

3.2. SILD

In many situations, determining the exact full class labels of the data samples is a hard task. Instead, collecting a kind of weak labels of the samples is easier to achieve. An example of this case include face verification under unconstrained conditions, where the only known information is whether pairs of face images belong the same persons or not. Such weak labels (pair

match/non match) are referred to as side information. In this case, the traditional LDA cannot work, since it relies on the class label, here the identity of the person, of each sample to compute the scatter matrices. To tackle this issue, Side-Information based Linear Discriminant Analysis (SILD) [17], which efficiently exploits the weak label information, has been proposed. In SILD, the positive pairs (match) are used to calculate the within scatter matrix S_{w} and the negative pairs (non match) are used to calculate the between scatter matrix S_{b}.

Let $\mathrm{P} = \{(\mathbf{u}_{\mathrm{i}}, \mathbf{u}_{\mathrm{j}}) : \mathrm{l}(\mathbf{u}_{\mathrm{i}}) = \mathrm{l}(\mathbf{u}_{\mathrm{j}})\}$ be the set of match images pairs and $\mathrm{Q} = \{(\mathbf{u}_{\mathrm{a}}, \mathbf{u}_{\mathrm{b}}) : \mathrm{l}(\mathbf{u}_{\mathrm{a}}) \neq \mathrm{l}(\mathbf{u}_{\mathrm{b}})\}$ be the set of non match images pairs, with l denotes the side information. The SILD within-class and between-class scatter matrices are estimated as follows:

$$S_{\mathrm{w}}^{\mathrm{sild}} = \sum_{(\mathbf{u}_{\mathrm{i}}, \mathbf{u}_{\mathrm{j}}) \in \mathrm{P}} (\mathbf{u}_{\mathrm{i}} - \mathbf{u}_{\mathrm{j}})(\mathbf{u}_{\mathrm{i}} - \mathbf{u}_{\mathrm{j}})^{\mathrm{T}} \tag{5}$$

$$S_{\mathrm{b}}^{\mathrm{sild}} = \sum_{(\mathbf{u}_{\mathrm{a}}, \mathbf{u}_{\mathrm{b}}) \in \mathrm{Q}} (\mathbf{u}_{\mathrm{a}} - \mathbf{u}_{\mathrm{b}})(\mathbf{u}_{\mathrm{a}} - \mathbf{u}_{\mathrm{b}})^{\mathrm{T}} \tag{6}$$

Thus, the scatter matrices computation do not require knowing the full class label of every sample. Instead, the side information is exploited to calculate the two matrices. Similarly to LDA, the projection matrix in SILD is obtained by solving the following optimization problem:

$$H_{\mathrm{opt}}^{\mathrm{sild}} = \underset{H}{\mathrm{argmax}} \frac{\left| H^{\mathrm{T}} S_{\mathrm{b}}^{\mathrm{sild}} H \right|}{\left| H^{\mathrm{T}} S_{\mathrm{w}}^{\mathrm{sild}} H \right|} \tag{7}$$

3.3. MSIDA

This part details the proposed extension of SILD to operate on multilinear data. Let $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_{\mathrm{N}}}$ be a tensor representation of a training samples. The set of training samples is organized into match tensor pairs $\mathrm{S_p} = \{(\boldsymbol{X}_{\mathrm{o}}, \boldsymbol{X}_{\mathrm{p}}) : \mathrm{l}(\boldsymbol{X}_{\mathrm{o}}) = \mathrm{l}(\boldsymbol{X}_{\mathrm{p}})\}$ and the set of non match tensor pairs $\mathrm{D_p} = \{(\boldsymbol{X}_{\mathrm{r}}, \boldsymbol{X}_{\mathrm{s}}) : \mathrm{l}(\boldsymbol{X}_{\mathrm{r}}) \neq (\boldsymbol{X}_{\mathrm{s}})\}$, where l here denotes the side information.

In order to enhance the discrimination among the tensors of different classes, a Discriminant Tensor Criterion (DTC) is imposed. In [15], the DTC ensures that the scatter of the intra-class samples is minimized and the scatter of the inter-class samples is maximized. Differently, in our case since we do not have the class labels, we minimize the scatter of the similar pairs and maximize the scatter of the dissimilar pairs, as done in SILD. The

DTC is achieved by multiple interrelated projections of the tensor through its different modes.Thus, the DTC in our case is given by:

$$U_{\mathrm{k}}^{*}\mid_{\mathrm{k=1}}^{\mathrm{N}}=\underset{U_{\mathrm{k}}\mid_{\mathrm{k=1}}^{\mathrm{N}}}{\operatorname{argmax}}\frac{\sum_{\mathrm{D_p}=1}^{\mathrm{n_{D_p}}}\parallel \boldsymbol{X}_{\mathrm{r}}\times_1 U_1\cdots\times_{\mathrm{N}} U_{\mathrm{N}}-\boldsymbol{X}_{\mathrm{s}}\times_1 U_1\cdots\times_{\mathrm{N}} U_{\mathrm{N}}\parallel^2}{\sum_{\mathrm{S_p}=1}^{\mathrm{n_{S_p}}}\parallel \boldsymbol{X}_{\mathrm{o}}\times_1 U_1\cdots\times_{\mathrm{N}} U_{\mathrm{N}}-\boldsymbol{X}_{\mathrm{p}}\times_1 U_1\cdots\times_{\mathrm{N}} U_{\mathrm{N}}\parallel^2} \tag{8}$$

$\mathrm{n_{S_p}}$ and $\mathrm{n_{D_p}}$ are number of positive pairs and negative pairs, respectively.

Equation (8) is an optimization problem with a high order nonlinear constraint. Consequently, finding its closed-form solution is a challenging task. An alternative solution is to use the iterative optimization approach as in [15] to estimate the interrelated projection matrices.

First, we consider one mode of the tensor, the optimization is formulated as:

$$U_{\mathrm{k}}^{*}=\underset{U_{\mathrm{k}}}{\operatorname{argmax}}\frac{\sum_{\mathrm{D_p}=1}^{\mathrm{n_{D_p}}}\parallel \boldsymbol{X}_{\mathrm{r}}\times_{\mathrm{k}} U_{\mathrm{k}}-\boldsymbol{X}_{\mathrm{s}}\times_{\mathrm{k}} U_{\mathrm{k}}\parallel^2}{\sum_{\mathrm{S_p}=1}^{\mathrm{n_{S_p}}}\parallel \boldsymbol{X}_{\mathrm{o}}\times_{\mathrm{k}} U_{\mathrm{k}}-\boldsymbol{X}_{\mathrm{p}}\times_{\mathrm{k}} U_{\mathrm{k}}\parallel^2} \tag{9}$$

As proven in [15], the optimization problem in equation (9) is a special case of discriminant analysis. To solve this problem, the tensors are unfolded into matrices in the k mode by equation (3). Then, the corresponding vectors of the unfolded matrix are attributed the same label as the original tensor and the positive vector pairs are used to compute the within-class scatter matrix and the negative vector pairs are used to compute the between-class scatter matrix. Thus optimization problem in Equation (9) could be rewrite as follows:

$$U_{\mathrm{k}}^{*}=\operatorname{argmax}_{U_{\mathrm{K}}}\frac{\mathrm{Tr}(\,U_{\mathrm{k}}^{\mathrm{T}}(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k}}U_{\mathrm{k}})}{\mathrm{Tr}(\,U_{\mathrm{k}}^{\mathrm{T}}(S^{\mathrm{msida}})_{\mathrm{w}}^{\mathrm{k}}U_{\mathrm{k}})} \tag{10}$$

where $(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k}}$ is the between-class scatter matrix for mode k calculated using the vectors of the negative pairs by:

$$(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k}}=\sum_{\mathrm{j=1}}^{\prod_{\mathrm{i}\neq\mathrm{k}}\mathrm{I_i}}(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k,j}},(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k,j}}=\sum_{(\boldsymbol{X}_{\mathrm{r}},\boldsymbol{X}_{\mathrm{s}})\in\mathrm{D_p}}(\mathbf{x}_{\mathrm{r}}^{\mathrm{k,j}}-\mathbf{x}_{\mathrm{s}}^{\mathrm{k,j}})(\mathbf{x}_{\mathrm{r}}^{\mathrm{k,j}}-\mathbf{x}_{\mathrm{s}}^{\mathrm{k,j}})^{\mathrm{T}} \tag{11}$$

similarly, in (10) $(S^{\text{msida}})_{\text{w}}^{\text{k}}$ is the within-class scatter matrix for mode k calculated using the corresponding vectors of unfolded matrix of the positive sample pairs:

$$(S^{\text{msida}})_{\text{w}}^{\text{k}} = \sum_{\text{j}=1}^{\prod_{\text{i}\neq\text{k}}\text{I}_\text{i}} (S^{\text{msida}})_{\text{w}}^{\text{k,j}}, (S^{\text{msida}})_{\text{w}}^{\text{k,j}} = \sum_{(\boldsymbol{X}_\text{o},\boldsymbol{X}_\text{p})\in\text{S}_\text{p}} (\mathbf{x}_\text{o}^{\text{k,j}} - \mathbf{x}_\text{p}^{\text{k,j}})(\mathbf{x}_\text{o}^{\text{k,j}} - \mathbf{x}_\text{p}^{\text{k,j}})^{\text{T}} \tag{12}$$

$\mathbf{x}^{\text{k,j}}$ denotes the j-th vector of the k mode unfolded matrix $X^{\text{k,j}}$ of tensor samples $\boldsymbol{X}$.

Now that the solution for one mode is known, the optimization of the entire tensor can be solved iteratively. The projection matrices $U_1, U_2, \cdots, U_\text{N}$ are first initialized to identity. At each iteration $U_1, \cdots, U_{\text{K}-1}, U_{\text{K}+1}, \cdots, U_\text{N}$ are supposed known and U_k is estimated. Setting $\boldsymbol{Y}_* = \boldsymbol{X}_* \times_1 U_1 \cdots \times_{\text{k}-1} U_{\text{k}-1} \times_{\text{k}+1} U_{\text{k}+1} \cdots \times_\text{N} U_\text{N}$, Equation (9) becomes:

$$U_\text{k}^* = \underset{U_\text{k}}{\text{argmax}} \frac{\sum_{\text{D}_\text{p}=1}^{\text{n}_{\text{D}_\text{p}}} \parallel \boldsymbol{Y}_\text{r} \times_\text{k} U_\text{k} - \boldsymbol{Y}_\text{s} \times_\text{k} U_\text{k} \parallel^2}{\sum_{\text{S}_\text{p}=1}^{\text{n}_{\text{S}_\text{p}}} \parallel \boldsymbol{Y}_\text{o} \times_\text{k} U_\text{k} - \boldsymbol{Y}_\text{o} \times_\text{k} U_\text{k} \parallel^2} \tag{13}$$

Equation (13) is similar to (10), which can be solved by eigen decomposition.

The iterative process of MSIDA breaks up on the realization of one of the following conditions: either i) the number of iterations reaches a predefined maximum; or ii) the norm of difference of the estimated projection between two successive iterations is below a threshold, $\left\|U_\text{k}^{\text{itr}} - U_\text{k}^{\text{itr}-1}\right\| \prec \text{I}_\text{k} \times \text{I}_\text{k} \times \epsilon$ where I_k is the K mode dimension and U_k^{itr} is the eigenvectors matrix in mode k calculated by:

$$(S^{\text{msida}})_{\text{b}}^{\text{k}} U_\text{k}^{\text{itr}} = \Lambda_\text{k} (S^{\text{msida}})_{\text{w}}^{\text{k}} U_\text{k}^{\text{itr}} \tag{14}$$

where Λ_k is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_\text{i}^\text{k}$.

The final lower dimensions $\acute{\text{I}}_1 \times \acute{\text{I}}_2 \cdots \times \acute{\text{I}}_\text{N}$ are set based on the percentage of energy (Energy_k) of eigenvalues to keep for each mode k:

$$\text{Energy}_\text{k} = \frac{\sum_{\text{i}=1}^{\acute{\text{I}}_\text{k}} \lambda_\text{i}^\text{k}}{\sum_{\text{i}=1}^{\text{I}_\text{k}} \lambda_\text{i}^\text{k}} \times 100 \tag{15}$$

with $\lambda_1^\text{k} > \lambda_2^\text{k} \cdots > \lambda_{\text{I}_\text{k}}^\text{k}$.

The entire procedure for the proposed Multilinear Side-Information based Discriminant Analysis (MSIDA) is provided in Algorithm 1.

4. MSIDA based face pair matching

In this section, we explain the detail of employing the proposed MSIDA for two different face matching applications: identity and kinship verification from pairs of face images. As depicted in Fig. 3, the block diagram of the proposed approach consists of three essential components: feature extraction, tensor subspace transformation and comparison.

Each face image is represented using two local descriptors, MSLPQ and MSBSIF, extracted at different scales yielding several feature vectors per face. The feature vectors of all training faces are arranged as a third order tensor $(\mathrm{i}_1, \mathrm{i}_2, \mathrm{i}_3)$, where i_1 corresponds to a single feature vector, i_2 corresponds the different local descriptors extracted at different scales, and i_3 corresponds to the face samples in the training database. This tensor is first projected using MPCA [13] to a lower subspace $\mathrm{j}_1 \times \mathrm{j}_2 \times \mathrm{i}_3$, where $\mathrm{j}_1 \times \mathrm{j}_2 << \mathrm{i}_2 \times \mathrm{i}_2$. The reason for applying MPCA prior to MSIDA is to avoid the small sample size problem in different tensor modes. This problem occurs when the dimension of features is larger than the number of samples, leading to the singularity of MSIDA scatter matrices. Reducing the dimension of each tensor mode first is therefore performed.

After applying MPCA, the training data tensor is split into two sub-tensors corresponding to the match pairs and non-match pairs, respectively. The split is done according the third mode i_3. The first tensor is used to compute the within-class scatter matrix $(S^{\mathrm{msida}})_{\mathrm{w}}$ and the second tensor is used to compute the between-class scatter matrix $(S^{\mathrm{msida}})_{\mathrm{b}}$ of the MSIDA. The data tensor is projected using MSIDA to a new lower and discriminative subspace $\mathrm{k}_1 \times \mathrm{k}_2$, where $\mathrm{k}_1 \times \mathrm{k}_2 << \mathrm{j}_1 \times \mathrm{j}_2$.

At the test phase, each of the face images of the pair to check the match is represented as a second order tensor formed by stacking the local descriptors of the image. Subsequently, both tensors are projected by MPCA then MSIDA. Finally, the cosine similarity between the pair is computed and used to check whether the pair is match (belonging to the same person/family) or not.

In the following, we explain the detailed steps of the proposed system.

Algorithm 1 Multilinear Side-Information based Discriminant Analysis

Input:

- Tensor $\widetilde{\boldsymbol{X}} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \cdots \times \mathrm{N} \times \mathrm{M}}$ of M training samples
- The weak labels ($\mathrm{labels}_{\mathrm{W}}$) for extracting the match tensor pairs $\mathrm{S_p} = \{(\boldsymbol{X}_{\mathrm{o}}, \boldsymbol{X}_{\mathrm{p}}) : \mathrm{l}(\boldsymbol{X}_{\mathrm{o}}) = \mathrm{l}(\boldsymbol{X}_{\mathrm{p}})\}$ and non match tensor pairs $\mathrm{D_p} = \{(\boldsymbol{X}_{\mathrm{r}}, \boldsymbol{X}_{\mathrm{s}}) : \mathrm{l}(\boldsymbol{X}_{\mathrm{r}}) \neq (\boldsymbol{X}_{\mathrm{s}})\}$ from $\widetilde{\boldsymbol{X}}$
- The maximal number of iterations: $\mathrm{Itr}_{\max}$
- The energy $(\mathrm{Energy}_1, \mathrm{Energy}_2, \cdots \mathrm{Energy}_{\mathrm{N}})$ used for selecting the final lower dimensions: $\acute{\mathrm{I}}_1 \times \acute{\mathrm{I}}_2 \cdots \times \acute{\mathrm{I}}_{\mathrm{N}}$.

Output:

- The different modes' projection matrices $U_{\mathrm{k}} = U_{\mathrm{k}}^{\mathrm{itr}} \in \mathbb{R}^{\mathrm{I_k} \times \acute{\mathrm{I}}_{\mathrm{k}}}, \mathrm{k} = 1, \cdots, \mathrm{N}$

Algorithm:

1. **Initialization:** $U_1^0 = I_{\mathrm{I}_1}, U_2^0 = I_{\mathrm{I}_2}, \cdots, U_{\mathrm{N}}^0 = I_{\mathrm{I_N}}$
2. **For** itr : 1 to $\mathrm{Itr}_{\max}$
   (a) **For** k=1 to N
   - $\boldsymbol{Y}_{\mathrm{i}} = \boldsymbol{X}_{\mathrm{i}} \times_1 U_1^{\mathrm{itr}-1} \cdots \times_{\mathrm{k}-1} U_{\mathrm{k}-1}^{\mathrm{itr}-1} \times_{\mathrm{k}+1} U_{\mathrm{k}+1}^{\mathrm{itr}-1} \cdots \times_{\mathrm{N}} U_{\mathrm{N}}^{\mathrm{itr}-1}$
   - $Y_{\mathrm{i}}^{\mathrm{k}} \Leftarrow_{\mathrm{k}} \boldsymbol{Y}_{\mathrm{i}}$
   - $(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k}} = \sum_{\mathrm{j}=1}^{\prod_{\mathrm{i} \neq \mathrm{k}} \mathrm{I_i}} (S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k,j}}, (S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k,j}} = \sum_{(\boldsymbol{Y}_{\mathrm{r}}, \boldsymbol{Y}_{\mathrm{s}}) \in \mathrm{D_p}} (\mathbf{y}_{\mathrm{r}}^{\mathrm{k,j}} - \mathbf{y}_{\mathrm{s}}^{\mathrm{k,j}})(\mathbf{y}_{\mathrm{r}}^{\mathrm{k,j}} - \mathbf{y}_{\mathrm{s}}^{\mathrm{k,j}})^{\mathrm{T}}$
   - $(S^{\mathrm{msida}})_{\mathrm{w}}^{\mathrm{k}} = \sum_{\mathrm{j}=1}^{\prod_{\mathrm{i} \neq \mathrm{k}} \mathrm{I_i}} (S^{\mathrm{msida}})_{\mathrm{w}}^{\mathrm{k,j}}, (S^{\mathrm{msida}})_{\mathrm{w}}^{\mathrm{k,j}} = \sum_{(\boldsymbol{Y}_{\mathrm{o}}, \boldsymbol{Y}_{\mathrm{p}}) \in \mathrm{S_p}} (\mathbf{y}_{\mathrm{o}}^{\mathrm{k,j}} - \mathbf{y}_{\mathrm{p}}^{\mathrm{k,j}})(\mathbf{y}_{\mathrm{o}}^{\mathrm{k,j}} - \mathbf{y}_{\mathrm{p}}^{\mathrm{k,j}})^{\mathrm{T}}$
   - Compute $(S^{\mathrm{msida}})_{\mathrm{b}}^{\mathrm{k}} U_{\mathrm{k}}^{\mathrm{itr}} = \Lambda_{\mathrm{k}} (S^{\mathrm{msida}})_{\mathrm{w}}^{\mathrm{k}} U_{\mathrm{k}}^{\mathrm{itr}}$

   (b) **If** itr >2 and $\left\| U_{\mathrm{k}}^{\mathrm{itr}} - U_{\mathrm{k}}^{\mathrm{itr}-1} \right\| < \mathrm{I_k} \times \mathrm{I_k} \times \epsilon, \mathrm{k} = 1, \cdots, \mathrm{N}$, break;
3. Compute the final lower dimensions $\acute{\mathrm{I}}_{\mathrm{k}}$ by: $\mathrm{Energy_k} = \frac{\sum_{\mathrm{i}=1}^{\acute{\mathrm{I}}_{\mathrm{k}}} \lambda_{\mathrm{i}}^{\mathrm{k}}}{\sum_{\mathrm{i}=1}^{\mathrm{I_k}} \lambda_{\mathrm{i}}^{\mathrm{k}}} \times 100, \mathrm{where}(\lambda_1^{\mathrm{k}} > \lambda_2^{\mathrm{k}} \cdots > \lambda_{\mathrm{I_k}}^{\mathrm{k}})$.
4. Sort the $\acute{\mathrm{I}}_{\mathrm{k}}$ eigenvectors $U_{\mathrm{k}} \in \mathbb{R}^{\mathrm{I_k} \times \acute{\mathrm{I}}_{\mathrm{k}}}$ according to $\lambda_{\mathrm{i}}^{\mathrm{k}}$ in decreasing order, $\mathrm{K} = 1, \cdots, \mathrm{N}$.

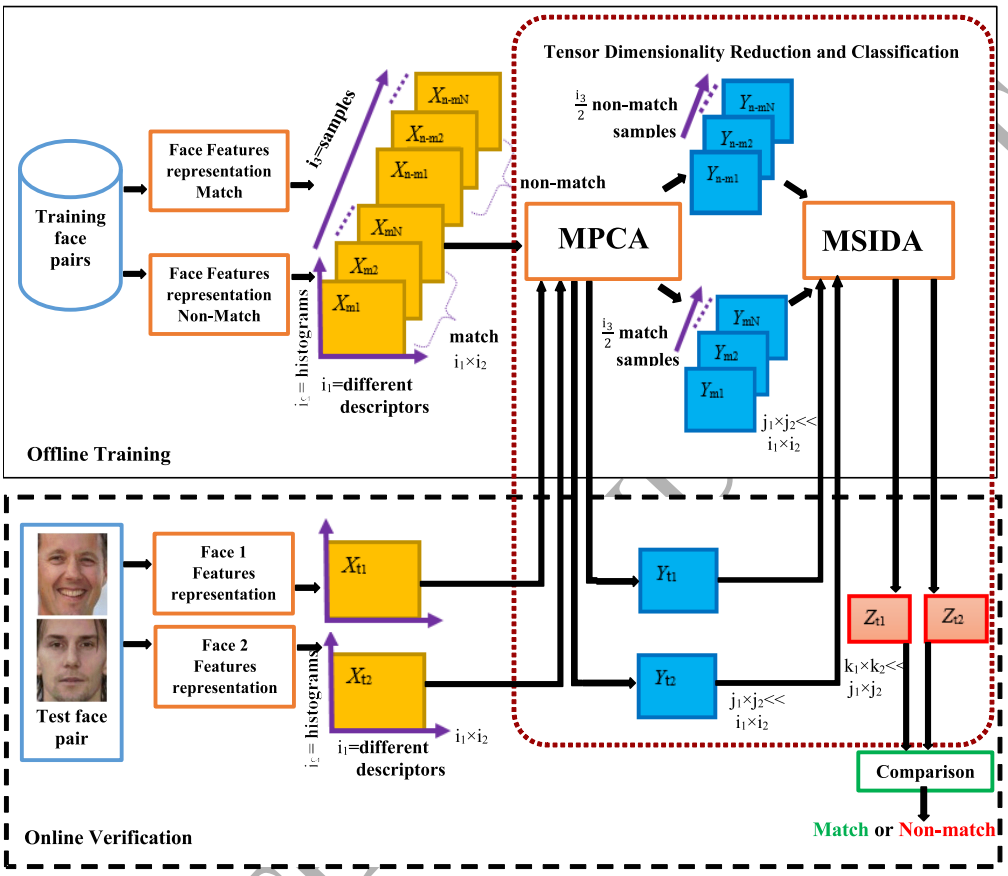


Figure 3: Block diagram of the proposed face pair matching system.

4.1. Features extraction

For feature extraction, we employ two local descriptors: local phase quantization (LPQ) [21] and binarized statistical image feature (BSIF) [22]. LPQ quantifies the phase of low frequencies while BSIF encodes the responses to pre-learned filters. We selected these two descriptors because they achieved the best performances in former works [23, 24, 25]. In order to keep the spatial face structure, the face image is subdivided into P non-overlapping face regions. The histograms of descriptors of the P regions are concatenated forming the face feature vector. To further enrich the face description, the two local descriptors are extracted at different scales. Fig. 4 depicts the

results of applying the two descriptors, at different scales, on a face image.

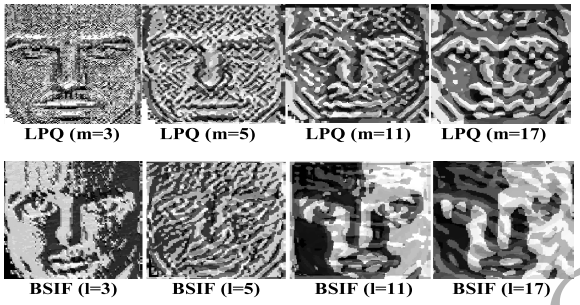


Figure 4: Face description using LPQ (top) and BSIF (bottom) at different scales.

For each face image, all the feature vectors of the two descriptors extracted at different scales are stacked in a matrix forming a second order tensor representation of the face. The faces in the database are arranged in a third order tensor and projected into a discriminative subspace tensor using MSIDA as explained in Section 3. Finally, the projected tensor is vectorized yielding the final face feature vector.

4.2. Matching

To compute the similarity between a face pair, we use the cosine similarity [26]. The use of cosine similarity measure is motivated by its link with Bayes decision rule ensuring the minimum classification error [27]. The cosine similarity between two vectors $\mathbf{z}_{t1}$ and $\mathbf{z}_{t2}$ is defined as follows:

$$\cos(\mathbf{z}_{t1}, \mathbf{z}_{t2}) = \frac{\mathbf{z}_{t1}^{T}.\mathbf{z}_{t2}}{\|\mathbf{z}_{t1}\|.\|\mathbf{z}_{t2}\|} \tag{16}$$

where $\|.\|$ is the Euclidean norm. A high value of the produced score means a high probability that the face pair belongs to the same person/family.

5. Experiments

In this section, we evaluate the proposed approach for two face matching problems, i.e. face based identity and kinship verification in the wild. We first introduce the four face datasets employed for experiments. Then, the parameter settings of the experiments are given and results are reported and discussed. Finally, we compare our best results with the state of art.

5.1. Datasets

Four different face datasets are considered for evaluating our approach. Following is a brief description of the data.

- **Labeled Faces in the Wild (LFW) dataset** [28] is a large face dataset collected from the web. LFW is particularly designed to study face recognition problem in unconstrained environments, covering real world variations in terms of pose, lighting, expressions, resolution, blur, occlusion, etc. This challenging dataset comprises 13,233 images from 5,749 different persons. The database is split into two views: view 1 used for model selection and view 2 used for performance evaluation. LFW defines three evaluation protocols: the image unrestricted setting, the image restricted setting and the unsupervised setting. In our experiments, we evaluate the proposed approach on the view 2 using aligned face images under image-restricted protocol, where no outside additional training data can be involved. The dataset is divided into 10 disjoint subsets for cross validation, where 9 subsets are used for training and the remaining subset for testing. Each subset contains 300 pairs of matched face pair and 300 pairs of mismatched face pairs. The performance is reported as the mean accuracy over the folds as well as the ROC curve over the 10-fold cross validation. Examples face images from this dataset are shown in Fig. 5.

Figure 5: Sample images from the LFW dataset. Left: positive pairs and right: negative pairs.

- **Cornell KinFace dataset** [5] comprises 150 pairs of face images of persons with a kin relation. The dataset was collected from the Internet considering four family relations. The family relations distribution in the dataset is 40% father son (F-S), 22% father-daughter (F-D), 13%, mother-son (M-S), and 25% mother-daughter (M-D). Each pair is composed of one parent face image and one child face image. In our

experiments, we evaluate the proposed approach on the same protocol as the one defined by Yan *et al.* [29], where only 143 pairs are used and classification is achieved with five-fold cross-validation for kin verification. The negative pairs are randomly generated by associating each parent image with an image of another child. Example face images from this dataset are shown in Fig. 6.

Figure 6: Sample images from the Cornell KinFace dataset. Left: positive pairs and right: negative pairs.

- **UB KinFace dataset** [30] comprises 600 images of 400 people, collected from the web. These images are separated into 200 families. Each family is composed of three images, which correspond to the child, young parent and old parent, respectively. This forms two subsets of pairs: set 1 of 200 child and 200 young parent and set 2 of 200 child and 200 old parent. In our experiments, we evaluate the proposed approach on the same protocol as defined by Yan *et al.* [29], using five-fold cross validation setting. Examples of face images from this dataset are shown in Fig. 7.

Figure 7: Sample images from the UB KinFace dataset. Left: positive pairs and right: negative pairs.

- **TSKinFace dataset** [31] is a newly published dataset, and it is among the largest available kinship databases. In total, the dataset counts

4060 face images. It includes two sets of three subject family relations: Father-Mother-Son (FM-S) and Father-Mother-Dauther (FM-D). There are 513 FM-S and 502 FM-D distinct groups of tri-subject. For our case, we restructured the dataset by splitting the set of Father-Mother-Son into two pairs Father-Son and Mother-Son ship and the set of Father-Mother-Daughter into two pairs Father-Daughter and Mother-Daughter. Following the same kinship verification protocol as defined by Qin *et al.* [31], we used five-fold cross-validation setting. Each fold comprise nearly the same number of face pairs. Examples of face images from this dataset are shown in Fig. 8.

Figure 8: Sample images from the TSKinFace dataset. Left: positive pairs and right: negative pairs.

5.2. Parameter settings

The parameters yielding the reported results of our experiments are provided in the following. The face images are cropped and resized to 130×90 pixels, 126×115 pixels, 96×80 pixels and 64×64 pixels for LFW dataset, Cornell, UB KinFace and TSKinface datasets, respectively. After extracting MSLPQ and MSBSIF from each face image, the resulting feature image is subdivided into 20 non-overlapping blocks. For extracting both descriptors at different scales, we vary the window size (m) of LPQ and the filter size (l) of BSIF. Same values are considered for both parameters m and l= $\{3, 5, 7, 9, 11, 13, 15, 17\}$. To keep the feature vector of reasonable size while maximizing the image description, the number of combined scales at the same time is varied from 2 to 4 for both descriptors. These parameters are adopted since their empirical results are proven the best.

The training data is utilized for estimating the subspace projection matrices. The MSIDA subspace is set to 96% energy of eigenvalues (as equation (15) indicates, the eigenvalues that save up to 96% of information are kept).

The number of iterations, for both MPCA and MSIDA algorithms, is empirically tuned and the better value is $\mathrm{Itr}_{\max} = 2$.

5.3. Results and discussion

This subsection presents the evaluation results and discussion. In addition to the third order tensor representation, all the experiments are also conducted for the original linear approach, SILD, which serves as a baseline for assessing our proposed method MSIDA. Furthermore, we examine the performances of both local descriptors, MSLPQ and MSBSIF, individually as well as their fusion. In the linear case, feature level fusion is done by concatenating vectors of different scales for each face descriptor. For proposed multilinear case, fusion is performed based on tensor, where vectors of different scales of two descriptors, LPQ and BSIF, are stacked in the second mode of the tensor.

To mitigate for the small sample size problem, feature dimensions are first reduced as follows. In linear (second order tensor) representation case, we apply Principal Component Analysis (PCA) prior to Side-Information based Linear Discriminant (SILD). In multilinear (third order tensor) representation case, we apply multilinear principal component analysis (MPCA) prior to Multilinear Side-Information based Discriminant Analysis (MSIDA).

Results, in terms of verification accuracy, are provided in Table 1 for face verification on LFW database and in Tables 2, 3 and 4 for kinship verification on Cornell KinFace, UB KinFace and TSKinFace datasets, respectively.

From the descriptor point of view and at individual descriptor level, we can notice that, in most of the cases, MSBSIF achieves better results than MSLPQ. we also can observe that the best results are obtained for MSBSIF with big-sized filters combination (13+15+17), whereas combination of medium-sized windows (5+9+13) of MSLPQ descriptors gives the best results. Furthermore, the fusion of features from the three previous filters of both descriptors permanently manifest better results across the four tables. Therefore, we select these two multiscale setting (13+15+17 for MSBSIF and 5+9+13 for MSLPQ) when fusing the two descriptors. This last fusion, achieves better results than the individual descriptors in all cases, demonstrating the complementary of MSBSIF and MSLPQ for face description.

Considering the influence of the subspace approaches, for each particular case of Tables 1, 2, 3 and 4, (regardless of the features, scales, databases, and face matching problem) the proposed multilinear approach MSIDA performs

better than its linear SILD counterpart. This clearly demonstrates the effectiveness of the proposed MSIDA. Deeper insights to the results reveal that MSIDA is able to extract better discriminative feature subspace that SILD. For instance, while in the linear case MSBSIF always performs better than MSLPQ, with perceivable margin, the difference in performance between the two descriptors is reduced in the multilinear case; and even there appear few cases where MSLPQ outperforms.

The results point out the fact that the weakly supervised multilinear method MSIDA takes full advantage of the higher order tensor structure. The curse of dimensionality dilemma is intuitively avoided. Indeed, while in the case of SILD method the feature vectors are simply concatenated ignoring the natural data structure, these features are stacked in a tensor mode allowing the extraction of maximum information. Furthermore, MSIDA abstracts better the discriminant information present in each mode compared with SILD method, which extracts discriminant information by a single projection.

Table 1: Verification accuracy of MSIDA and SILD using different MSLPQ and MSBSIF scales and their fusion on the LFW dataset under restricted protocol.

Feature	Mean Accuracy (%) ± Standard Error	
	PCA+SILD	MPCA+MSIDA
MSLPQ (m=3+5)	89.60 ± 1.21	93.07 ± 1.06
MSLPQ (m=5+9+13)	90.43 ± 1.25	94.33 ± 0.94
MSLPQ (m=13+15+17)	90.33 ± 1.21	94.02 ± 1.03
MSLPQ (m=5+7+9+11)	90.03 ± 1.28	93.57 ± 1.04
MSBSIF (l=3+5)	90.27 ± 1.39	92.27 ± 1.34
MSBSIF(l=5+9+13)	89.00 ± 1.29	94.05 ± 0.95
MSBSIF (l=13+15+17)	90.47 ± 1.21	94.37 ± 0.85
MSBSIF (l=5+7+9+11)	89.60 ± 1.11	93.03 ± 1.14
MSLPQ (m=5+9+13) + MSBSIF (l=13+15+17)	**91.30 ± 1.26**	**94.40 ± 0.89**

5.4. Effect of the quantity of training pairs

We investigate the effect of the amount of training pairs for LFW, Cornell KinFace, UB KinFace and TSKinFace datasets. Different amounts of randomly selected training image pairs are used in this experiment. Specifically, 600, 1200, 1800, 2400, 3000, 3600, 4200, 4800 and 5400 pairs are used for

Table 2: Verification accuracy of MSIDA and SILD using different MSLPQ and MSBSIF scales and their fusion on the Cornell KinFace dataset.

Feature	Mean Accuracy (%)	
	PCA+SILD	MPCA+MSIDA
MSLPQ (m=3+5)	78.97	83.67
MSLPQ (m=5+9+13)	81.77	86.11
MSLPQ (m=13+15+17)	76.90	85.78
MSLPQ (m=5+7+9+11)	78.64	85.80
MSBSIF (l=3+5)	83.21	84.67
MSBSIF(l=5+9+13)	82.52	85.00
MSBSIF (l=13+15+17)	83.23	86.45
MSBSIF (l=5+7+9+11)	82.92	84.67
MSLPQ (m=5+9+13) + MSBSIF (l=13+15+17)	**84.56**	**86.59**

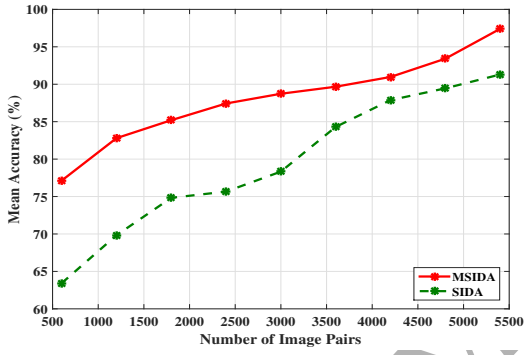
LFW database. For Cornell and UB KinFace databases, we use 30, 60, 90, 120, 150, 180 and 210 pairs. For TSKinFace, we use 80, 160, 240, 320, 400, 480, 560, 640, 720 and 800 pairs. Fig. 9 shows the mean accuracy of the proposed method MSIDA and its counterpart SILD with respect to the number of training pairs. We can clearly observe that the performances of the both methods MSIDA and SILD become better as the amount of training pairs increase. Moreover, our method MSIDA outperforms SILD on all datasets for different amounts of training data. The experimental results indicate that the quantity of training pairs has a significant influence on the final accuracy. This is exhibited by the accuracy improvements between any two successive configurations of the experiment.

5.5. Weakly supervised MSIDA against full supervised LDA and MDA

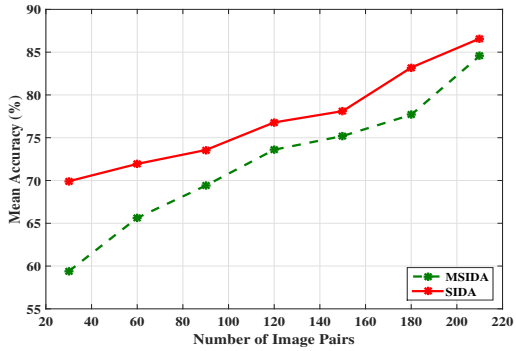
In this experiment, we compare the basic SILD and the proposed method MSIDA under the restricted setting with LDA and MDA under the unrestricted setting on the LFW dataset. The verification accuracy of the different methods MSIDA, SILD, MDA, and LDA under restricted and unrestricted settings on the LFW dataset are shown in Table 5, their ROC curves are illustrated in Fig 10 . We can observe that MSIDA and SILD provide comparable performances to their full supervised counterparts, MDA and LDA. Also, the margin of the improvement of MDA in the presence of identity labels is smaller compared to the side-information setting with MSIDA

Figure 9: Mean accuracy of SILD and MSIDA with different number of pairs on (a) LFW (b) Cornell KinFace (c) UB KinFace (d) TSKinFace databases, respectively.

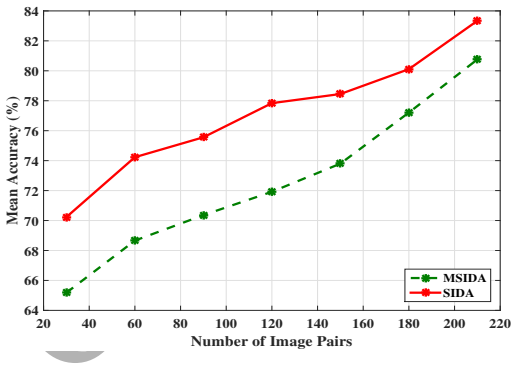
(a)



(b)



(c)



(d)

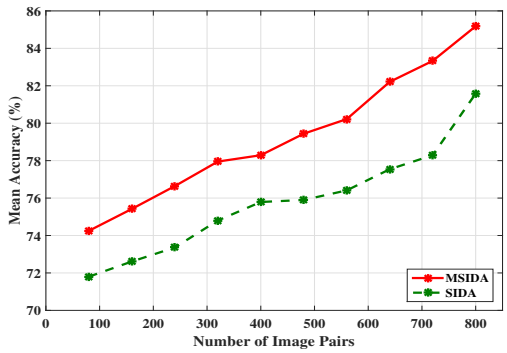


Table 3: Verification accuracy of MSIDA and SILD using different MSLPQ and MSBSIF scales and their fusion on the UB KinFace dataset.

Feature	Mean Accuracy (%)	
	PCA+SILD	MPCA+MSIDA
MSLPQ (m=3+5)	78.35	82.82
MSLPQ (m=5+9+13)	78.89	82.97
MSLPQ (m=13+15+17)	76.03	82.56
MSLPQ (m=5+7+9+11)	78.77	82.21
MSBSIF (l=3+5)	79.08	82.84
MSBSIF(l=5+9+13)	78.99	82.98
MSBSIF (l=13+15+17)	79.30	83.14
MSBSIF (l=5+7+9+11)	79.13	82.81
MSLPQ (m=5+9+13) + MSBSIF (l=13+15+17)	**80.76**	**83.34**

(only a small difference of less than 0.67%). Therefore, we can confirm that MSIDA with weakly information achieves comparable performances when using the full class label in the supervised case with MDA. In addition, with weakly labeled information, our method MSIDA is better than LDA with full labeled information. This advantage presents the proposed MSIDA as an effective model to deal with the weakly information problem.

5.6. Time complexity

The experiments were done using MATLAB version 2015a on a PC with an Intel Xeon(R) 3.19 GHz CPU and 12 GB of RAM. The estimation of the projection matrices is implemented in the offline training phase. The complexity in terms of computation time (CT) for one face pair matching (during the online phase) using SILD and the proposed MSIDA is provided in Table 6. Since the same features (MSLPQ+MSBSIF) are used to describe the face image for both SILD and MSIDA, the same CT is required for both methods in this step. For the dimensionality reduction and comparison, it can be seen that our method MSIDA achieves a good gain in the CT compared with the baseline method in all datasets. The CT is reduced thanks to the use of tensor representation of face data. In addition to accuracy, this experiment demonstrates the effectiveness of our approach in terms of computation complexity. Indeed, the MSIDA does not exceed a total CT of

Table 4: Verification accuracy of MSIDA and SILD using different MSLPQ and MSBSIF scales and their fusion on the TSKinFace dataset.

Feature	Mean Accuracy (%)	
	PCA+SILD	MPCA+MSIDA
MSLPQ (m=3+5)	78.20	82.61
MSLPQ (m=5+9+13)	79.94	84.22
MSLPQ (m=13+15+17)	79.51	83.98
MSLPQ (m=5+7+9+11)	79.90	83.78
MSBSIF (l=3+5)	76.67	79.60
MSBSIF (l=5+9+13)	78.20	80.80
MSBSIF (l=13+15+17)	79.99	83.24
MSBSIF (l=5+7+9+11)	79.53	80.80
MSLPQ (m=5+9+13) + MSBSIF (l=13+15+17)	**81.58**	**85.18**

Table 5: Verification accuracy (%) of the different methods under restricted and unrestricted settings on the LFW dataset.

Setting	weakly supervised (without full class label)		full Supervised (with full class label)	
	Restricted		Unrestricted	
Method	MSIDA	SILD	MDA	LDA
Mean Accuracy (%) ± Standard Error (%)	94.40 0.89	91.30 ± 1.26	**95.07± 0.71**	91.93 ± 0.82

0.1259 s, which makes it a competent candidate for use in real world face verification applications.

Table 6: Time complexity in seconds of MSIDA and SILD on the four datasets.

Dataset	Feature extraction	Dimensionality reduction and comparison		Total run time	
		SILD	MSIDA	SILD	MSIDA
LFW	0.1085	0.0217	0.0174	0.1302	0.1259
Cornell KinFace	0.1249	0.0045	0.0032	0.1294	0.1281
UB KinFace	0.0978	0.0050	0.0036	0.1028	0.1014
TSKinface	0.0785	0.0041	0.0029	0.0826	0.0814

5.7. Comparison with state of the art

5.7.1. Face verification in the Wild

Our comparison criteria is the accuracy obtained by the recent works not only on the side of similar mechanisms.The comparison of our approach

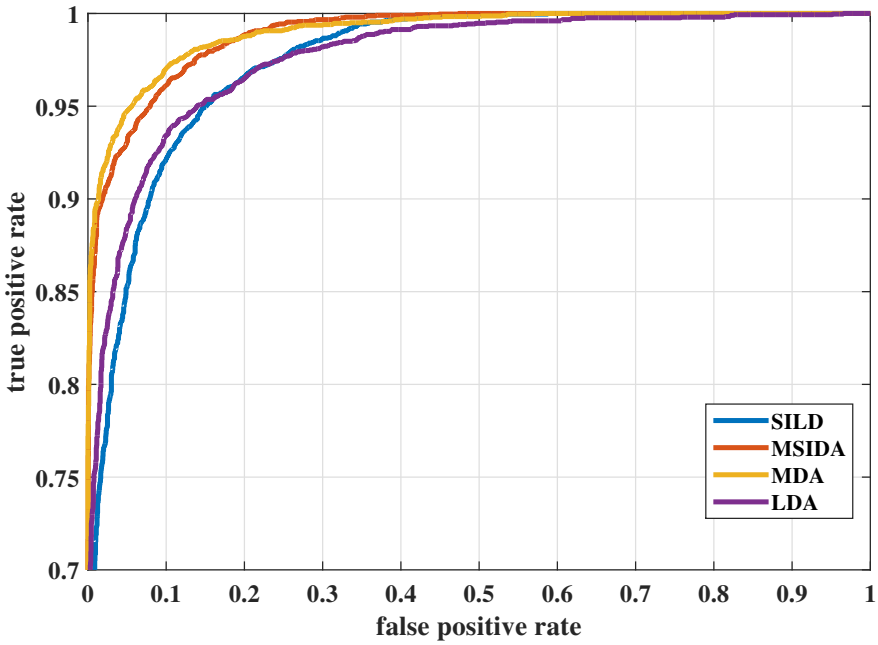


Figure 10: ROC curves of the different methods under restricted and unrestricted settings on the LFW dataset.

on the side of metric learning and feature type is provided on the tables 1, 2, 3, 4 and 5. In this subsection, table 7 compares our results with the state-of-the-art including the latest works on the LFW database under restricted setting protocol. Additionally, Fig. 11 depicts the corresponding ROC curves . The best verification accuracy of our work is 94.40%. This result is ranked secon of different methods, among the state of the art. The recent work MRF-Fusion-CSKDA [38] yields the currently best verification accuracy on the LFW dataset. MRF-Fusion-CSKDA is based on the fusion of three multiscale descriptors MSLBP, MSLPQ and MSBSIF using kernel methods. In contrast, we employed only two descriptors, MSLPQ and MSBSIF using the proposed multilinear approaches MSIDA achieving second rank in the litterature. Hence, our system is less complex, computationally efficient and scales better than the top ranking method.

Table 7: Comparison of verification accuracy of MSIDA with state of the art on the LFW dataset under restricted protocol.

Method	Mean Accuracy (%) ± Standard Error (%)
Eigenfaces, original [32]	60.02 ± 0.79
Nowak2, funneled [33]	73.93 ± 0.49
Hybrid descriptor-based, funneled [34]	78.47 ± 0.51
Pixels/MKL, funneled [35]	68.22 ± 0.41
Fisher vector faces [36]	87.47 ± 1.49
Eigen-PEP [37]	88.97 ± 1.32
class-specific kernel discriminant analysis-MRF-MSLPQ [38]	92.42 ± 1.03
class-specific kernel discriminant analysis-MRF-MSBSIF [38]	93.63 ± 1.27
class-specific kernel discriminant analysis MRF-Fusion [38]	95.89 ± 1.94
Hierarchical-PEP [39]	91.10 ± 1.47
Side-Information based Linear Discriminant Analysis -MSBSIF [24]	90.37 ± 1.19
Robust Statistical Frontalization [40]	88.81 ± 0.78
Discriminative Deep Multi-Metric Learning [41]	93.28 ± 0.39
MSIDA-MLPQ	94.33 ± 0.94
MSIDA-MBSIF	94.37 ± 0.85
MSIDA-fusion (MLPQ+MBSIF)	**94.40 ± 0.89**

5.7.2. Kinship Verification in the wild

Table 8, Table 9 and Table 10 compare the proposed approach with the state-of-the-art on the Cornell KinFace, UB KinFace and TSKinFace datasets, respectively. We note that the best verification accuracies is obtained by our approach ain which we have 86.87% on Cornell KinFace, 83.34% on UB KinFace and 85.18% on TSKinFace. As can be seen from the three comparison tables, our approach outperforms the other state-of-the-art methods on the three kinship datasets. Moreover, we can clearly see that MSIDA improves the kinship performance with a significant margin (For instance, more than 13% on Cornell KinFace and more than 11% improvement on UB KinFace).Since our feature learning approach is similar to the several approaches, the power of our approach is given thanks to the multilinear sceme based on the proposed MSIDA using high order tensor representation that exploit more discriminative information.

6. Conclusion

In this paper, we presented an effective approach for matching images of face pairs in the wild. The proposed approach is based on high order tensor representation of face images. The face tensor is built using histograms of two

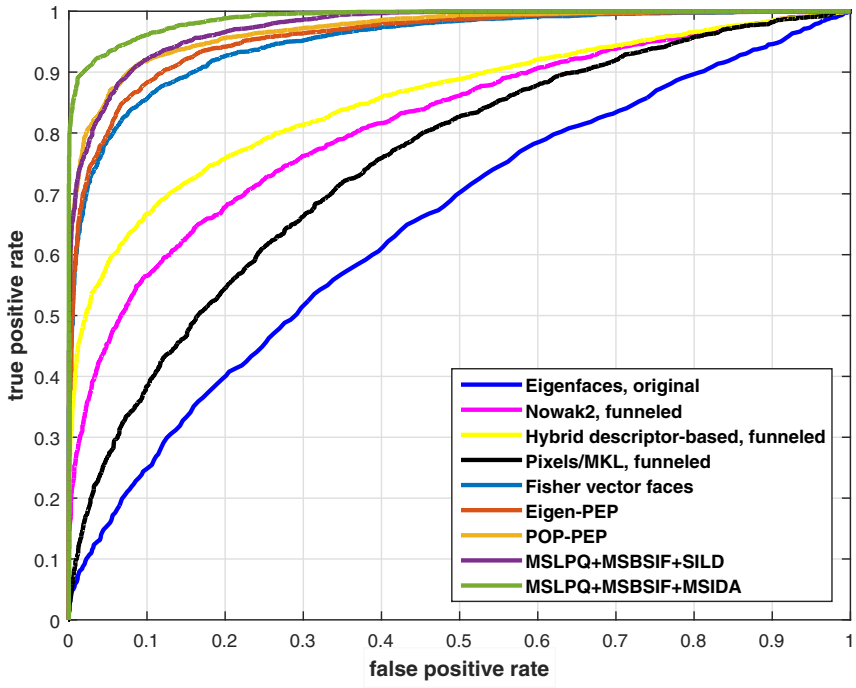


Figure 11: ROC curve of MSIDA and state-of-the-art on the LFW dataset under image restricted protocol.

local descriptors, MBSIF and MLPQ, extracted at multiscales. To project the face tensor into a low dimensional and discriminative subspace, taking advantage of the available weakly labeled data, we have extended the linear SILD to MSIDA, which operates on multilinear data. MSIDA finds multilinear projections of the tensor, where the separation between data classes is enhanced exploiting the available side information. The experimental evaluation of our approach for identity and kinship verification demonstrates that the proposed method is more effective, both in terms of accuracy and computation time, than its original linear form. Additionally, our approach outperforms the state of the art on three kinship databases (Cornell KinFace, UB KinFace, and TSKinFace) databases and ranks second on LFW database. As a future direction, we plan to investigate other multilinear dimensionality reduction methods and examine higher tensor orders (> 3) for face representation.

Table 8: Comparison of verification accuracy of MSIDA with state of the art on the Cornell KinFace dataset.

Method	Mean Accuracy (%)
Pictorial structure model-HOG [5]	70.67
Discriminative multimetric learning-fusion (LBP, SIFT, SPLE) [29]	73.50
Multiview Neighborhood Repulsed Metric Learning-fusion (LBP, SIFT, TPLBP and LE) [6]	71.60
Prototype discriminative feature learning-fusion (LBP, SIFT) [42]	71.90
Multiple Kernel Similarity Metric-LPQ [43]	81.70
MSIDA-MLPQ	86.11
MSIDA-MBSIF	86.45
MSIDA-fusion (MLPQ+MBSIF)	**86.87**

Table 9: Comparison of verification accuracy of MSIDA with state of the art on the UB KinFace dataset.

Method	Mean Accuracy (%)
Transfer subspace learning-Gabor [44]	56.50
Self-similarity representation of Weber faces-DOG [45]	53.90
Discriminative multimetric learning-fusion (LBP, SIFT, SPLE)	72.25
Multiview Neighborhood Repulsed Metric Learning-fusion (LBP, SIFT, TPLBP and LE) [6]	67.05
Prototype discriminative feature learning-fusion (LBP, SIFT) [42]	67.30
Random Subset Feature Selection-fusion (LPQ, LBP, SPLE) [46]	76.65
MSIDA-MLPQ	82.97
MSIDA-MBSIF	83.14
MSIDA-fusion (MLPQ+MBSIF)	**83.34**

Table 10: Comparison of verification accuracy of MSIDA with state of the art on the TSKinFace dataset.

Method	Mean Accuracy (%)
Relative Symmetric Bilinear Model-SIFT [31]	81.85
Binarized Statistical Image Feature [47]	74.46
Discriminative Deep Multi-Metric Learning-LPQ [41]	79.92
Multiple Kernel Similarity Metric-LPQ [43]	81.89
MSIDA-MLPQ	83.98
MSIDA-MBSIF	83.24
MSIDA-fusion (MLPQ+MBSIF)	**85.18**

References

References

[1] L. Best-Rowden, S. Bisht, J. C. Klontz, A. K. Jain, Unconstrained face recognition: Establishing baseline human performance via crowdsourcing, in: IEEE International Joint Conference on Biometrics, 2014, pp. 1–8. doi:10.1109/BTAS.2014.6996296.

[2] M. D. Marsico, M. Nappi, D. Riccio, H. Wechsler, Robust face recognition for uncontrolled pose and illumination changes, IEEE Transactions on Systems, Man, and Cybernetics: Systems 43 (1) (2013) 149–163. doi:10.1109/TSMCA.2012.2192427.

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

[4] G. B. Huang, H. Lee, E. Learned-Miller, Learning hierarchical representations for face verification with convolutional deep belief networks, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2518–2525. doi:10.1109/CVPR.2012.6247968.

[5] R. Fang, K. D. Tang, N. Snavely, T. Chen, Towards computational models of kinship verification, in: 2010 IEEE International Conference on Image Processing, 2010, pp. 1577–1580. doi:10.1109/ICIP.2010.5652590.

[6] J. Lu, X. Zhou, Y. P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (2) (2014) 331–345. doi:10.1109/TPAMI.2013.134.

[7] N. Kohli, M. Vatsa, R. Singh, A. Noore, A. Majumdar, Hierarchical representation learning for kinship verification, IEEE Transactions on Image Processing 26 (1) (2017) 289–302. doi:10.1109/TIP.2016.2609811.

[8] S. J. Wang, S. Yan, J. Yang, C. G. Zhou, X. Fu, A general exponential framework for dimensionality reduction, IEEE Transactions on Image Processing 23 (2) (2014) 920–930. doi:10.1109/TIP.2013.2297020.

[9] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang, B. Xu, Generalized discriminant analysis: A matrix exponential approach, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40 (1) (2010) 186–197. doi:10.1109/TSMCB.2009.2024759.

[10] S. Liu, Q. Ruan, C. Wang, G. An, Tensor rank one differential graph preserving analysis for facial expression recognition, Image Vision Comput. 30 (8) (2012) 535–545. doi:10.1016/j.imavis.2012.05.004. URL http://dx.doi.org/10.1016/j.imavis.2012.05.004

[11] Y. Pang, S. Wang, Y. Yuan, Learning regularized lda by clustering, IEEE Transactions on Neural Networks and Learning Systems 25 (12) (2014) 2191–2201. doi:10.1109/TNNLS.2014.2306844.

[12] Z. Lai, Y. Xu, J. Yang, J. Tang, D. Zhang, Sparse tensor discriminant analysis, IEEE Transactions on Image Processing 22 (10) (2013) 3904–3915. doi:10.1109/TIP.2013.2264678.

[13] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Mpca: Multilinear principal component analysis of tensor objects, IEEE Transactions on Neural Networks 19 (1) (2008) 18–39. doi:10.1109/TNN.2007.901277.

[14] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition, IEEE Transactions on Neural Networks 20 (1) (2009) 103–123. doi:10.1109/TNN.2008.2004625.

[15] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H. J. Zhang, Multilinear discriminant analysis for face recognition, IEEE Transactions on Image Processing 16 (1) (2007) 212–220. doi:10.1109/TIP.2006.884929.

[16] M. Safayani, M. T. M. Shalmani, Three-dimensional modular discriminant analysis (3dmda): A new feature extraction approach for face recognition, Computers & Electrical Engineering 37 (5) (2011) 811–823.

[17] M. Kan, S. Shan, D. Xu, X. Chen, Side-information based linear discriminant analysis for face recognition, in: British Machine Vision Conference, BMVC 2011, Dundee, UK, August 29 - September 2, 2011. Proceedings, 2011, pp. 1–12. doi:10.5244/C.25.125.
URL http://dx.doi.org/10.5244/C.25.125

[18] A. Ouamane, M. Bengherabi, A. Hadid, M. Cheriet, 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, 2015, pp. 1–6. doi:10.1109/FG.2015.7284837.

[19] C. H. Chan, J. Kittler, N. Poh, T. Ahonen, M. Pietikinen, (multiscale) local phase quantisation histogram discriminant analysis with score normalisation for robust face recognition, in: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, 2009, pp. 633–640. doi:10.1109/ICCVW.2009.5457642.

[20] S. R. Arashloo, Multiscale binarised statistical image features for symmetric unconstrained face matching, in: 2014 22nd Iranian Conference on Electrical Engineering (ICEE), 2014, pp. 1377–1382. doi:10.1109/IranianCEE.2014.6999748.

[21] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: Proceedings of the 3rd International Conference on Image and Signal Processing, ICISP '08, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 236–243. doi:10.1007/978-3-540-69905-7_27.
URL http://dx.doi.org/10.1007/978-3-540-69905-7_27

[22] J. Kannala, E. Rahtu, Bsif: Binarized statistical image features., in: ICPR, IEEE Computer Society, 2012, pp. 1363–1366.
URL http://dblp.uni-trier.de/db/conf/icpr/icpr2012.html#KannalaR12

[23] A. Ouamane, B. Messaoud, A. Guessoum, A. Hadid, M. Cheriet, Multi scale multi descriptor local binary features and exponential discriminant analysis for robust face authentication, in: 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 313–317. doi:10.1109/ICIP.2014.7025062.

[24] A. Ouamane, M. Bengherabi, A. Hadid, M. Cheriet, 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, 2015, pp. 1–6. doi:10.1109/FG.2015.7284837.

[25] A. Chouchane, M. Belahcène, A. Ouamane, S. Bourennane, Evaluation of Histograms Local Features and Dimensionality Reduction for 3D Face Verification, Journal of Information Processing Systems (2016) 1–21. URL https://hal.archives-ouvertes.fr/hal-01281110

[26] H. V. Nguyen, L. Bai, Cosine Similarity Metric Learning for Face Verification, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 709–720. doi:10.1007/978-3-642-19309-5_55.
URL http://dx.doi.org/10.1007/978-3-642-19309-5_55

[27] C. Liu, Discriminant analysis and similarity measure, Pattern Recogn. 47 (1) (2014) 359–367. doi:10.1016/j.patcog.2013.06.023.
URL https://doi.org/10.1016/j.patcog.2013.06.023

[28] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments, in: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Marseille, France, 2008.
URL https://hal.inria.fr/inria-00321923

[29] H. Yan, J. Lu, W. Deng, X. Zhou, Discriminative multimetric learning for kinship verification, IEEE Transactions on Information Forensics and Security 9 (7) (2014) 1169–1178. doi:10.1109/TIFS.2014.2327757.

[30] S. Xia, M. Shao, J. Luo, Y. Fu, Understanding kin relationships in a photo, IEEE Transactions on Multimedia 14 (4) (2012) 1046–1056. doi:10.1109/TMM.2012.2187436.

[31] X. Qin, X. Tan, S. Chen, Tri-subject kinship verification: Understanding the core of A family, CoRR abs/1501.02555. arXiv:1501.02555.
URL http://arxiv.org/abs/1501.02555

[32] M. A. Turk, A. P. Pentland, Face recognition using eigenfaces, in: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision

and Pattern Recognition, 1991, pp. 586–591. doi:10.1109/CVPR.1991.139758.

[33] G. B. Huang, V. Jain, E. Learned-Miller, Unsupervised joint alignment of complex images, in: 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8. doi:10.1109/ICCV.2007.4408858.

[34] L. Wolf, T. Hassner, Y. Taigman, Descriptor based methods in the wild, in: Real-Life Images workshop at the European Conference on Computer Vision (ECCV), 2008.
URL http://www.openu.ac.il/home/hassner/projects/Patchlbp

[35] N. Pinto, J. J. DiCarlo, D. D. Cox, 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, 2009, pp. 2591–2598. doi:10.1109/CVPR.2009.5206605.

[36] K. Simonyan, O. M. Parkhi, A. Vedaldi, A. Zisserman, Fisher vector faces in the wild, in: BMVC, 2013.

[37] H. Li, G. Hua, X. Shen, Z. Lin, J. Brandt, Eigen-PEP for Video Face Recognition, Springer International Publishing, Cham, 2015, pp. 17–33. doi:10.1007/978-3-319-16811-1_2.
URL http://dx.doi.org/10.1007/978-3-319-16811-1_2

[38] S. R. Arashloo, J. Kittler, Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features, IEEE Transactions on Information Forensics and Security 9 (12) (2014) 2100–2109. doi:10.1109/TIFS.2014.2359587.

[39] H. Li, G. Hua, Hierarchical-pep model for real-world face recognition, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4055–4064. doi:10.1109/CVPR.2015.7299032.

[40] C. Sagonas, Y. Panagakis, S. Zafeiriou, M. Pantic, Robust statistical frontalization of human and animal faces, International Journal of Computer Vision 122 (2) (2017) 270–291. doi:10.1007/s11263-016-0920-7.
URL https://doi.org/10.1007/s11263-016-0920-7

[41] J. Lu, J. Hu, Y. P. Tan, Discriminative deep metric learning for face and kinship verification, IEEE Transactions on Image Processing 26 (9) (2017) 4269–4282. doi:10.1109/TIP.2017.2717505.

[42] H. Yan, J. Lu, X. Zhou, Prototype-based discriminative feature learning for kinship verification, IEEE Transactions on Cybernetics 45 (11) (2015) 2535–2545. doi:10.1109/TCYB.2014.2376934.

[43] Y.-G. Zhao, Z. Song, F. Zheng, L. Shao, Learning a multiple kernel similarity metric for kinship verification, Information Sciences 430-431 (2018) 247 – 260. doi:https://doi.org/10.1016/j.ins.2017.11.048.
URL http://www.sciencedirect.com/science/article/pii/S0020025516310416

[44] M. Shao, S. Xia, Y. Fu, Genealogical face recognition based on ub kinface database, in: CVPR 2011 WORKSHOPS, 2011, pp. 60–65. doi:10.1109/CVPRW.2011.5981801.

[45] N. Kohli, R. Singh, M. Vatsa, Self-similarity representation of weber faces for kinship classification, in: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2012, pp. 245–250. doi:10.1109/BTAS.2012.6374584.

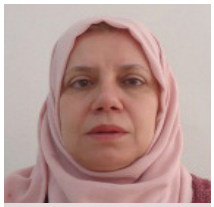
[46] P. Alirezazadeh, A. Fathi, F. Abdali-Mohammadi, Effect of purposeful feature extraction in high-dimensional kinship verification problem 3 (2018) 183–191.

[47] X. Wu, E. Boutellaa, M. B. Lpez, X. Feng, A. Hadid, On the usefulness of color for kinship verification from face images, in: 2016 IEEE International Workshop on Information Forensics and Security (WIFS), 2016, pp. 1–6. doi:10.1109/WIFS.2016.7823901.

Biography

Mohcene Bessaoudi He has a master degree in Electronic Telecommunication from University of Mohammed Khider Biskra, Algeria in 2015 at the Department of Electrical Engineering. Currently, he is a Ph.D. student at the same university. His research interests are in: image processing, feature extraction, face detection, tensor analysis, classification, computer vision, biometric techniques and face recognition systems.

Abdelmalik Ouamane received the Doctor of Science degree from University of Mohammed Khider Biskra, Algeria in 2015. Currently, he is a Lecturer at Electrical Engineering Department at the same university. He has been a reviewer for many conferences and journals. His research interests include image processing, classification, multimodal biometric and tensor analysis.



Mebarka Belahcene received her Ph.D degree from University of Mohammed Khider Biskra, Algeria in 2013. She is currently a Lecturer at Electrical Engineering Department at the same university. Her research interests include signal processing, image processing and biometrics.

Ammar Chouchane received the Ph.D degree in Electronic Telecommunication from University of Mohammed Khider Biskra, Algeria in 2016. His research interests are in image processing, feature extraction, face detection, tensor analysis, classification, computer vision, biometric techniques and face recognition systems.

Elhocine Boutellaa received Ph.D. degree in Computer Science from Ecole Nationale Supérieure dInformatique, Algeria. He has been working as research associate with Centre de Développement des Technologies Avancées since 2010. Currently he holds the position of postdoctoral researcher at University of Oulu. His research interests include face analysis, biometrics and computer vision.

Salah Bourennane received his Ph.D. degree from Institut National Polytechnique de Grenoble, France, in 1990 in signal processing. Currently, he is a full Professor at the Ecole Centrale de Marseille, France. His research interests are in statistical signal processing, array processing, image processing, tensor signal processing, and performances analysis.