Knowledge-based tensor subspace analysis system for kinship verification

I. Serraoui, O. Laiadi, A. Ouamane, F. Dornaika, Abdelmalik Taleb-Ahmed

Knowledge-based Tensor Subspace Analysis System for Kinship Verification

I. Serraoui[b,d], O. Laiadi[c,d], A. Ouamane[b], F. Dornaika[a,e,f,*], A. Taleb-Ahmed[d,**]

[a] *Henan University, Kaifeng, China*
[b] *Laboratory of LI3C, University of Biskra, Algeria*
[c] *Laboratory of LESIA, University of Biskra, Algeria*
[d] *Univ. Polytechnique Hauts-de-France, CNRS, Univ. Lille, ISEN, Centrale Lille, UMR 8520 - IEMN, F-59313 Valenciennes, France*
[e] *University of the Basque Country UPV/EHU, San Sebastian, Spain*
[f] *IKERBASQUE Foundation for Science, Bilbao, Spain*

Abstract

Most existing automatic kinship verification methods focus on learning the optimal distance metrics between family members. However, learning facial features and kinship features simultaneously may cause the proposed models to be too weak. In this work, we explore the possibility of bridging this gap by developing knowledge-based tensor models based on pre-trained multi-view models. We propose an effective knowledge-based tensor similarity extraction framework for automatic facial kinship verification using four pre-trained networks (i.e., VGG-Face, VGG-F, VGG-M, and VGG-S). Therefore, knowledge-based deep face and general features (such as identity, age, gender, ethnicity, expression, lighting, pose, contour, edges, corners, shape, etc.) were successfully fused by our tensor design to understand the kinship cue. Multiple effective representations are learned for kinship verification statements (children and parents) using a margin maximization learning scheme based on Tensor Cross-view Quadratic Exponential Discriminant Analysis. Through the exponential learning process, the large gap between distributions of the same family can be reduced to the maximum, while the small gap between distributions of different families is si-

*Corresponding author: fadi.dornaika@ehu.eus
**Corresponding author: Abdelmalik.Taleb-Ahmed@uphf.fr

multaneously increased. The WCCN metric successfully reduces the intra-class variability problem caused by deep features. The explanation of black-box models and the problems of ubiquitous face recognition are considered in our system. The extensive experiments on four challenging datasets show that our system performs very well compared to state-of-the-art approaches.
Keywords: kinship verification, knowledge-based tensor subspace analysis, convolutional neural networks, multi-view deep features, metric learning, facial images analysis.

---

1. Introduction

The task of kinship verification aims to determine whether two peoples are biologically related based on their appearance [1]. Image-based kinship verification allows learning and extracting similarities between family members, which is a new challenge in face image analysis. This presents a new challenge for face image analysis. Moreover, this can provide valuable clues for many potential applications such as searching for missing children and human trafficking, organizing family albums, creating family trees, forensics, and image labeling. There were challenges in automatic kinship recognition using facial images such as different age groups and genders. Also, the images were taken in uncontrolled environments and without constraints on expression, partial occlusion, lighting, and background.

Kin verification schemes typically utilize shallow features-based face representation to learn the facial features and kinship traits simultaneously [1, 2]. This results in weakly learned kinship verification models. Several methods used a hierarchical approach to learn kinship features. First, they used knowledge-based pre-trained models that incorporate a better understanding the facial cues, and then project this knowledge to learn kinship cues from understandable faces of family members. Therefore, frameworks that used metric learning over knowledge-based deep features achieved the best results [3, 4, 5, 6, 7, 8, 9].

Tensor-based metric learning methods (multi-linear subspace) show better

performance than vector-based (linear subspace) counterparts using shallow features [10, 11] for kinship verification topic. Face feature extraction is an important step for tensor subspace analysis methods. The authors in [12] discuss in their paper and conclude that the face contains heterogeneous information where the face contains large dissimilarity parts (i.e. eyes, nose, mouth, etc.). It follows that the tensor-based methods perform better when using texture information (shallow features) and that the raw face images are poor input information for the tensor-based methods. Many works [13, 10, 11] propose the use of histograms of multiple local texture descriptors as an elegant method for describing face images. This confirms that texture representation (shallow representation) is more discriminative for tensor based face images.

However, deep features show great improvement over shallow features for computer vision topics such as facial expression classification [14], age estimation [15], gender classification [16], and pain estimation [17].

Explaining black-box models and ubiquitous face recognition are challenging topics to address with deep face recognition models. Based on the perspectives in [18, 19], a popular approach uses simpler and more axiomatic decision algorithms to build an alternative black-box model (e.g., a deep-learning algorithm) that is used for decision making. However, there is a risk that the alternative paradigm is too large to be truly comprehensible to humans.

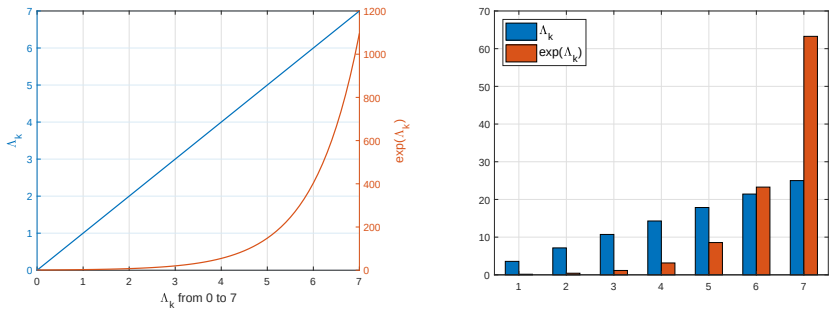
Many of previous studies [3, 4, 5, 7, 9] aims to project deep features from the unknown subspaces (black-box subspaces) to the known and more discriminative subspaces (e.g., microaggregation and shallow decision trees [20]). Through this transformation, they benefit from the knowledge-based domain and their proposed frameworks guarantee the transparency, efficiency and robustness of the proposed models based on deep features.

The authors in [21] conclude in their review for face recognition that the solution is to build a generic model trained on generic propositions and then applied to a specific application (in our case, a facial kinship verification application) and consolidate multiple sources of biometric data. Therefore, deep features obtained from different pre-trained networks trained on multiple state-

ments of biometric evidence are considered as a rich source of information that can be compatible to different biometric attributes.

Dornaika et al. [4] propose for the first time to use deep object information about contour, edges, corners, shape, etc. of facial parts (i.e., mouth, nose, eyebrows, eyes, etc.) alongside deep facial features (i.e., identity, age, gender, ethnicity, expression, illumination, pose, etc.) for the topic of kinship verification by fusing the VGG-F [22] and VGG-Face [23] features using the Multi-view Neighborhood Repulsed Metric Learning [1] subspace (linear subspace).

In general, metric learning methods focus on increasing the spread between samples of different classes while minimizing the spread between samples of the same class. One of the most popular criteria is the exponential criterion [24, 25, 26]. Figure 1 shows the exponential process at different eigenvalues. Through this application, the metric learning methods can greatly extend and improve their separation (classification). Thus, the eigenvalues with lower importance were minimized and those with high importance were maximized.



(a) (b)

Figure 1: Explanation of (a) Linear and exponential functions applied on different eigenvalues, and (b) Example of the proportions $\frac{\Lambda_k}{\sum \Lambda_k}$ (blue bars) and $\frac{exp(\Lambda_k)}{\sum exp(\Lambda_k)}$ (orange bars).

Tensor metric learning methods have yet to show an advantage in using deep features for facial kinship verification, where deep object and face features seem to be more promising. Moreover, the design of tensor subspace analysis (multi-linear subspace) needs to be increasingly adapted to the deep

features. Cross-view metric learning methods [11, 27] show great improvement over traditional metric learning in computer vision. For this reason, we aim to fully exploit the dependency structure between parents and children in a tensor cross-view manner. We call our method Tensor Cross-view Quadratic Exponential Discriminant Analysis integrating Within-Class Covariance Normalization (TXQEDA+WCCN).

The obtained results confirm that the use of deep object features (using a network trained on object classification) are a good complement to deep face features by tensor cross-view learning and the proposed ensemble can realize, fuse, analyse and separate the multi-factorial structure (extracted from deep face and object features) of face images related to identity, age, gender, ethnicity, expression, illumination, pose, contour, edges, corners, shape, etc over the various dimensions of tensor design [28]. Moreover, the VGG-Face model achieved the best performance in face recognition with accuracy of 98.95% and 97.3% on LFW [29] and YTF [30] databases, respectively. Our knowledge-based tensor subspace model also addresses the problem of ubiquitous face recognition by fusing features extracted from the VGG-Face model with those from the VGG-F, VGG-M, and VGG-S models via a tensor design. Moreover, we succeed in transferring the one entry problem to the model (i.e., face recognition/object detection) to deal with the binary and ternary entries model (bi-subject and tri-subject kinship verification). Moreover, our TXQEDA+WCCN approach addresses the problems of black-box models and considers the following properties: accuracy, fidelity, consistency, stability, representativeness, certainty, novelty, importance, and comprehensibility. These need to be considered in kinship verification frameworks. The main idea of our proposed knowledge-based framework for kinship verification problem can be explained in the following points:

1. We derive a novel hybrid knowledge-based multi-linear subspace for the kinship verification task. The proposed kinship verification subspace is based on multi-view deep networks to study the efficiency of different deep face image representations. While feature fusion has already been

exploited in each tensor mode unfolding.

2. We propose an improvement scheme based on margin maximization using both cross-view similarity learning and exponential discriminant analysis.

3. Large intra-class variance is considered to be the main challenge in kinship verification tasks. Inspired by this problem, we applied WCCN to solve the intra-class variability problem. The combination of TXQEDA and WCCN would minimize the expected classification error.

4. Many experiments were conducted with four public databases (TSKinFace, Cornell, KinFaceW-I, and KinFaceW- II), with favorable results compared to state-of-the-art approaches

The remainder of this paper is organized as follows. We begin with an overview of our system methodology in Section 2. In Section 3, we give an overview of our knowledge-based multi-linear subspace for face kinship verification. The TXQEDA+WCCN approach is presented in Section 4. The experimental data, setup and results are presented in Section 5. Finally, conclusions are drawn in Section 6.

2. Methodology

As mentioned earlier, the main idea of our work is to develop an effective metric learning that provides a small margin for points with kin relation and a large margin for points with no kin relation over a tensor subspace for kinship verification. More specifically, we propose the following novel criterion: Tensor Cross-view Quadratic Exponential Discriminant Analysis integrating Within-Class Covariance Normalization (TXQEDA+WCCN) method. For brevity, we denote a particular image representation, we extract two layers of features, namely, the fully connected layer 6 and 7 (FC6 and FC7) through VGG-F, VGG-M, VGG-S [22] and VGG-Face [23] pre-trained models to train the proposed TXQEDA+WCCN scheme. The overview of the proposed knowledge-based system is shown in Figure 2. For the offline training phase we extracted FC6

and FC7 layers using four deep feature descriptors (i.e., VGG-F, VGG-M, VGG-S, and VGG-Face).

The feature vectors of all training faces of parents sub-set or children sub-set are expressed by a 3^{rd} order tensor $\boldsymbol{A}, \boldsymbol{B} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ (shown in Figure 3), where $\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3$ represent the weights, the fully connected layers (FC6 and FC7) of each descriptor, and the samples respectively. The proposed TXQEDA+WCCN projected and reduced the input tensors $\boldsymbol{A}, \boldsymbol{B}$ corresponding to I_1 and I_2 modes. Consequently, we obtain a reduced tensor with $\mathrm{i}_1 \times \mathrm{i}_2 << \mathrm{I}_1 \times \mathrm{I}_2$. TXQEDA+WCCN estimates the projection matrices $W_{\mathrm{k}}^{\mathrm{iter}} \in \Re^{\mathrm{I_k} \times \mathrm{i_k}}$ for each k^{th} mode tensor.

In the online verification phase (test phase), the feature extraction for the test inputs (parent-child test pair) was performed in the same way as in the training phase, where each test image is represented as a tensor of order 2^{nd}. Then, both test pair tensors are projected by TXQEDA+WCCN, then we obtain two output matrices E_1 and E_2 with dimensions i_1, i_2 where $\mathrm{i}_1 \times \mathrm{i}_2 << \mathrm{I}_1 \times \mathrm{I}_2$. These two matrices ($E_1$ and E_2) are concatenated to a feature vector $\mathbf{e}_1$ and $\mathbf{e}_2$ of length $\mathrm{i}_1 \times \mathrm{i}_2$. Finally, to compute the similarity between the test pairs we use the cosine similarity between the two vectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

The Receiver Operating Characteristic (ROC) curve compares the cosine score to an iterative threshold during performance evaluation. A high value of the ROC curve score implies that there is a high probability that the test pairs belongs to the same family.

3. Knowledge-based multi-linear subspace

It is well known that strong facial features are crucial in determining and verifying the relationship between people. Therefore, the deep features are represented to explain kinship attribution becomes an important issue and a major challenge in determining kinship. Recently, deep features show better performance than shallow features (i.e. LPQ, LBP, BSIF ...etc.) in kinship verification. To find efficient and discriminative features from face images, We suggest taking advantage of prior knowledge and fuse the deep features provided

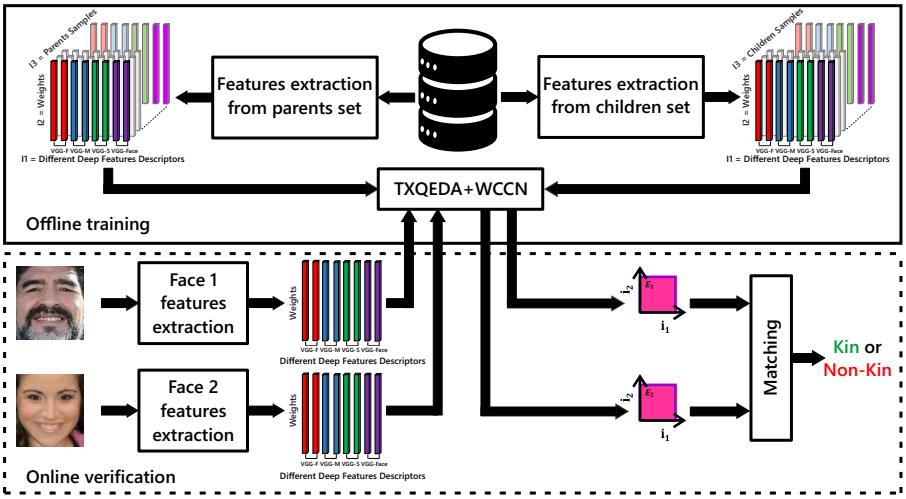


Figure 2: The proposed knowledge-based tensor subspace analysis system.

by four pre-trained networks VGG-F, VGG-M, VGG-S and VGG-Face to extract a complementary information to solve the kin relations between samples in multi-linear subspace and generate the kinship factor information.

The proposed multi-linear subspace transfers the knowledge from large-scale data-driven face and object recognition to the kinship verification system to find a metric space in which the facial features of family members become discriminative via a multi-linear cross-view. Figure 3 illustrates the proposed knowledge-based multi-linear subspace, each face image is represented using four deep feature descriptors (i.e., VGG-F, VGG-M, VGG-S, and VGG-Face) extracted in two fully connected layers FC6 and FC7, each with 4096 neurons. Thus, the feature vectors are considered as 2D arrays for each face. The features of the face images in each database are arranged to form a third-order tensor. Furthermore, feature fusion has already been brought to unfold in each tensor mode. Our method concerns the feature fusion process without using the usual feature fusion techniques. In other words, using a multi-dimensional tensor-based metric learning method simplifies the feature fusion task.

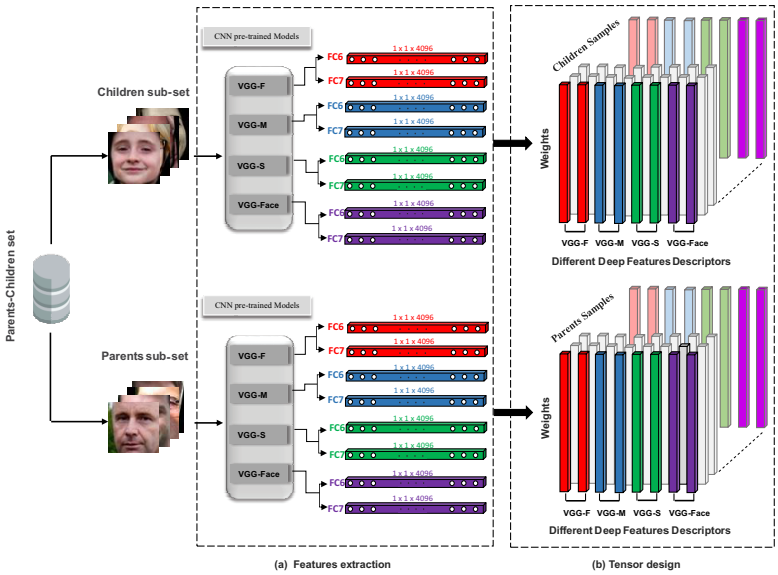


Figure 3: Knowledge-based multi-view deep features extraction (Mv-VGG) and tensor design.

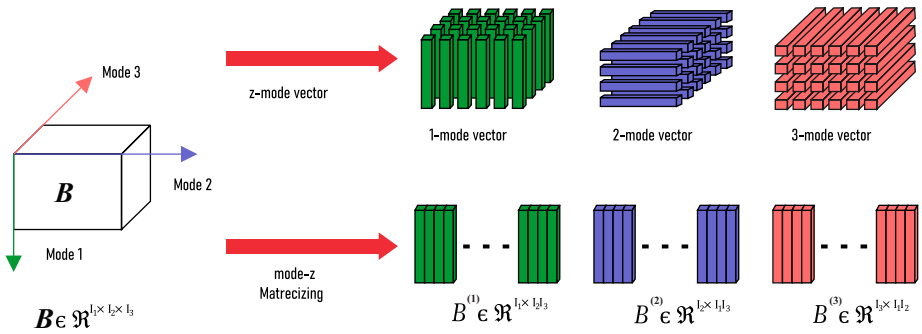


Figure 4: Visualisation of 3D subset unfolding, where Mode 1, Mode 2 and Mode 3 define weights, scales and samples, receptively.

Figure 4 illustrates how unfolding is used to perform deep feature fusion. The process of unfolding is performed mathematically as follows:

For a tensor $\boldsymbol{B} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$ with size $(I_1 \times I_2 \times \cdots \times I_N)$ tensor, then element $(i_1 \times i_2 \times \cdots \times i_N)$ of $\boldsymbol{B}$ maps to element (i_k, j) of matrix $B^{(k)} \in \Re^{I_k \times \prod_{i \neq k} I_i}$, the k-mode unfolding of $\boldsymbol{B}$, with:

$$B^k \Leftarrow_k \boldsymbol{B} \tag{1}$$

Where

$$B^{(k)}_{i_k,j} = \boldsymbol{B}_{i_1 \ldots i_N}, \quad j = 1 + \sum_{l=1, l\neq k}^{N} (i_l - 1) \prod_{o=l+1, o\neq k}^{N} I_o \tag{2}$$

For mode 3, we ignored the unfolding process, since this dimension contains the number of samples, while the learning of the kinship verification is based on these samples. Therefore, we avoid the unfolding process in the third mode

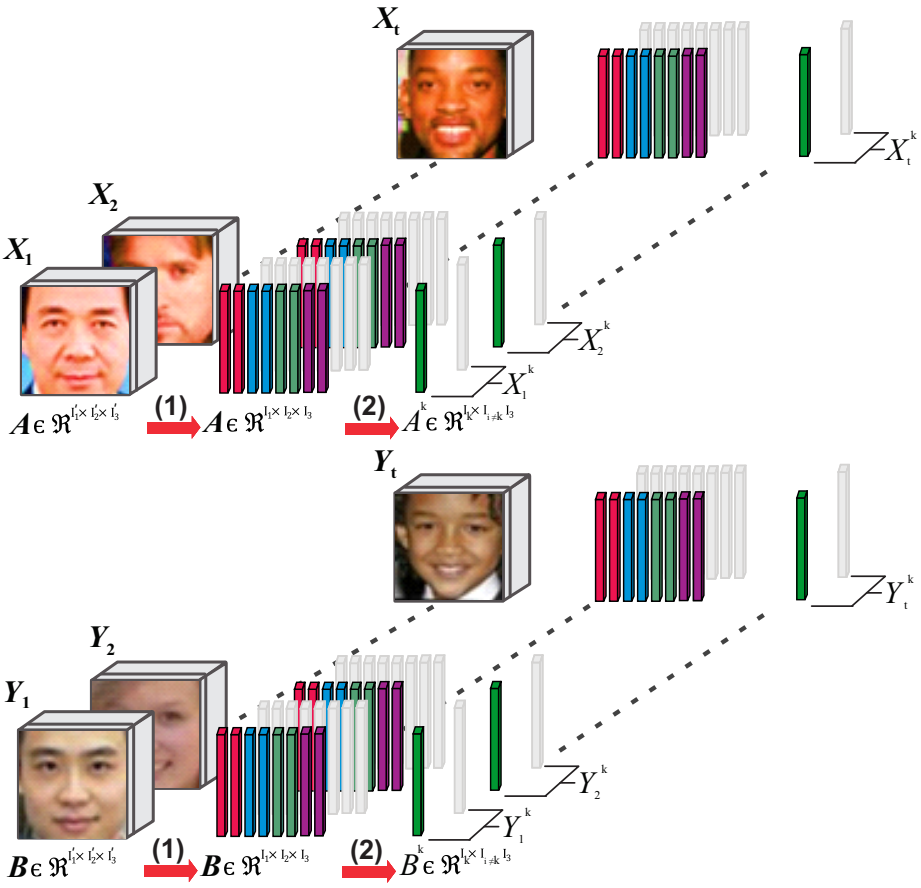


Figure 5: An example of the parent/child subset $\{\boldsymbol{A},\boldsymbol{B}\}$ in the case of each class (family) contains more than one face image for each individual. The first transition **(1)** is from raw images to deep features, while the second transition **(2)** is from tensor representation of deep features to unfolding representation.

4. Tensor Cross-view Quadratic Exponential Discriminant Analysis Integrating Within Class Covariance Normalization

This section focuses on the algebra of multi-linear Cross-view Quadratic Exponential Discriminant Analysis (TXQEDA) involving the Within Class Covariance Normalization metric (WCCN). We define a specific symbolic notation for different data types and use bold italic uppercase letters, normal italic uppercase letters, bold lowercase letters, lowercase letters and uppercase letters to denote tensors, matrices, vectors and scalars respectively.

4.1. Tensor Cross-view Quadratic Exponential Discriminant Analysis (TXQEDA)

In particular, given samples from training set $\{\boldsymbol{A}, \boldsymbol{B}\}$ of t classes/families with n/m images. Since each image provides 8 deep feature vectors, the input 3D tensors for parents and children are as follows. $\boldsymbol{A} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ $(\mathrm{I}_1 = 8, \mathrm{I}_2 = 4096, \mathrm{I}_3 = \mathrm{n})$ is parents sub-set contains t individuals $\boldsymbol{A} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_\mathrm{t}\}$ (the first view), and $\boldsymbol{B} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ $(\mathrm{I}_1 = 8, \mathrm{I}_2 = 4096, \mathrm{I}_3 = \mathrm{m})$ is children sub-set contains t individuals $\boldsymbol{B} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_\mathrm{t}\}$ (the second view), where $\boldsymbol{X}_\mathrm{i} = \{\mathrm{p}_1, \mathrm{p}_2, \ldots, \mathrm{p}_{\mathrm{n}_\mathrm{i}}\}$ and $\boldsymbol{Y}_\mathrm{i} = \{\mathrm{c}_1, \mathrm{c}_2, \ldots, \mathrm{c}_{\mathrm{m}_\mathrm{i}}\}$ $\mathrm{i} \in [1, \mathrm{t}]$, each individual from parents or children $\boldsymbol{X}_\mathrm{i}$, $\boldsymbol{Y}_\mathrm{i}$ contains $\mathrm{n_i}, \mathrm{m_i}$ images. Then, the training samples are transformed into a new unfolded subspace using Equations (1) and (2). We obtain new training sets $A^k \in \Re^{\mathrm{I_k} \times \mathrm{I}_{\mathrm{i} \neq \mathrm{k}} \mathrm{I}_3}$ and $B^k \in \Re^{\mathrm{I_k} \times \mathrm{I}_{\mathrm{i} \neq \mathrm{k}} \mathrm{I}_3}$ (see Figure 5 for further illustrations of the data structure).

We have:

$A^\mathrm{k} = \{X_1^\mathrm{k}, X_2^\mathrm{k}, \ldots, X_\mathrm{t}^\mathrm{k}\}$, $X_\mathrm{i}^\mathrm{k} = \{\mathbf{p}_1^\mathrm{k}, \mathbf{p}_2^\mathrm{k}, \ldots, \mathbf{p}_{\mathrm{n_i}}^\mathrm{k}\}$ $\mathrm{i} \in [1, \mathrm{t}]$

$B^\mathrm{k} = \{Y_1^\mathrm{k}, Y_2^\mathrm{k}, \ldots, Y_\mathrm{t}^\mathrm{k}\}$, $Y_\mathrm{i}^\mathrm{k} = \{\mathbf{c}_1^\mathrm{k}, \mathbf{c}_2^\mathrm{k}, \ldots, \mathbf{c}_{\mathrm{m_i}}^\mathrm{k}\}$ $\mathrm{i} \in [1, \mathrm{t}]$

Laiadi et al. [11], extended XQDA [27] method into TXQDA [11], the TXQDA objective is to estimate N projection matrices $(W_1 \in \Re^{\mathrm{I}_1 \times \mathrm{i}_1}, W_2 \in \Re^{\mathrm{I}_2 \times \mathrm{i}_2}, \ldots, W_\mathrm{N} \in \Re^{\mathrm{I_N} \times \mathrm{i_N}})$, for each k^{th}, mode tensor, which has a form of Generalized Rayleigh Quotient:

$$J(W_\mathrm{k}) = \frac{\mathrm{Tr}\left(W_\mathrm{k}^T \Sigma_\mathrm{E}^\mathrm{k} W_\mathrm{k}\right)}{\mathrm{Tr}\left(W_\mathrm{k}^T \Sigma_\mathrm{I}^\mathrm{k} W_\mathrm{k}\right)} \tag{3}$$

Besides the two cross-view matrices $\Sigma_{\mathrm{I}}^{\mathrm{k}}$ and $\Sigma_{\mathrm{E}}^{\mathrm{k}}$ can be efficiently calculated without actually computing the n/m pairwise differences, by simplifying them as follows:

$$\begin{aligned} \mathrm{n_I}\Sigma_{\mathrm{I}}^{\mathrm{k}} &= \sum_{\mathrm{f}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_I}\Sigma_{\mathrm{I}}^{\mathrm{k,f}} \\ \mathrm{n_I}\Sigma_{\mathrm{I}}^{\mathrm{k,f}} &= \tilde{P}^{\mathrm{k,f}}(\tilde{P}^{\mathrm{k,f}})^T + \tilde{C}^{\mathrm{k,f}}(\tilde{C}^{\mathrm{k,f}})^T - S^{\mathrm{k,f}}(R^{\mathrm{k,f}})^T - R^{\mathrm{k,f}}(S^{\mathrm{k,f}})^T \end{aligned} \tag{4}$$

Where:

$\tilde{P}^{\mathrm{k,f}} = (\sqrt{\mathrm{m_1}}X_1^{\mathrm{k,f}}, \sqrt{\mathrm{m_2}}X_2^{\mathrm{k,f}}, \ldots, \sqrt{\mathrm{m_t}}X_{\mathrm{t}}^{\mathrm{k,f}})$

$\tilde{C}^{\mathrm{k,f}} = (\sqrt{\mathrm{n_1}}\,Y_1^{\mathrm{k,f}}, \sqrt{\mathrm{n_2}}\,Y_2^{\mathrm{k,f}}, \ldots, \sqrt{\mathrm{n_t}}\,Y_{\mathrm{t}}^{\mathrm{k,f}})$

$\tilde{S}^{\mathrm{k,f}} = (\sum_{\mathrm{y_i}=1,\mathrm{j}=1}^{\mathrm{n_1}} \mathbf{p}_{\mathrm{j}}^{\mathrm{k,f}}, \sum_{\mathrm{y_i}=2,\mathrm{j}=1}^{\mathrm{n_2}} \mathbf{p}_{\mathrm{j}}^{\mathrm{k,f}}, \ldots, \sum_{\mathrm{y_i}=\mathrm{t},\mathrm{j}=1}^{\mathrm{n_t}} \mathbf{p}_{\mathrm{j}}^{\mathrm{k,f}})$

$\tilde{R}^{\mathrm{k,f}} = (\sum_{\mathrm{l_i}=1,\mathrm{j}=1}^{\mathrm{m_1}} \mathbf{c}_{\mathrm{j}}^{\mathrm{k,f}}, \sum_{\mathrm{l_i}=2,\mathrm{j}=1}^{\mathrm{m_2}} \mathbf{c}_{\mathrm{j}}^{\mathrm{k,f}}, \ldots, \sum_{\mathrm{l_i}=\mathrm{t},\mathrm{j}=1}^{\mathrm{m_t}} \mathbf{c}_{\mathrm{j}}^{\mathrm{k,f}})$

$\mathrm{n_I} = \sum_{\mathrm{i}=1}^{\mathrm{t}} \mathrm{n_i} \times \mathrm{m_i}$

For all representations, $\mathrm{y_i}$ and $\mathrm{l_j}$ are class/family labels, $\mathrm{n_i}$ is the number of images for each $\mathrm{i_{th}}$ individual X_{i}, and $\mathrm{m_i}$ is the number of images for each $\mathrm{i^{th}}$ individual Y_{i}.

$$\begin{aligned} \mathrm{n_E}\Sigma_{\mathrm{E}}^{\mathrm{k}} &= \sum_{\mathrm{f}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_E}\Sigma_{\mathrm{E}}^{\mathrm{k,f}} \\ \mathrm{n_E}\Sigma_{\mathrm{E}}^{\mathrm{k,f}} &= \mathrm{m}P^{\mathrm{k,f}}(P^{\mathrm{k,f}})^T + \mathrm{n}C^{\mathrm{k,f}}(C^{\mathrm{k,f}})^T - \mathbf{s}^{\mathrm{k,f}}(\mathbf{r}^{\mathrm{k,f}})^T - \mathbf{r}^{\mathrm{k,f}}(\mathbf{s}^{\mathrm{k,f}})^T - \mathrm{n_I}\Sigma_{\mathrm{I}}^{\mathrm{k,f}} \end{aligned} \tag{5}$$

Where :

$P^{\mathrm{k,f}} = (X_1^{\mathrm{k,f}}, X_2^{\mathrm{k,f}}, \ldots, X_{\mathrm{t}}^{\mathrm{k,f}})$

$C^{\mathrm{k,f}} = (Y_1^{\mathrm{k,f}}, Y_2^{\mathrm{k,f}}, \ldots, Y_{\mathrm{t}}^{\mathrm{k,f}})$

$\mathbf{s}^{\mathrm{k,f}} = \sum_{\mathrm{j}=1}^{\mathrm{t}} \sum_{\mathrm{i}=1}^{\mathrm{n_j}} \mathbf{p}_{\mathrm{i}}^{\mathrm{k,f}}$ and $\mathbf{r}^{\mathrm{k,f}} = \sum_{\mathrm{j}=1}^{\mathrm{t}} \sum_{\mathrm{j}=1}^{\mathrm{m_j}} \mathbf{c}_{\mathrm{j}}^{\mathrm{k,f}}$

$\mathrm{n_E} = \mathrm{n_p} \times \mathrm{n_c} - \mathrm{n_I}$ where, $\mathrm{n_p}$ and $\mathrm{n_c}$ represent the number of parents and children images.

It can be observed that TXQEDA computes Σ_{I} and Σ_{E} directly from the samples mean and covariance of each class/family and all classes/families (samples of parents from $\boldsymbol{A}$ and samples of children from $\boldsymbol{B}$, i.e. only positive pairs

in training data). Therefore, the matrices , $\tilde{P}^{\mathrm{k,f}}$, $S^{\mathrm{k,f}}$, $P^{\mathrm{k,f}}$ $s^{\mathrm{k,f}}$ depend on data samples from the parents' view, while the matrices $\tilde{C}^{\mathrm{k,f}}$, $R^{\mathrm{k,f}}$, $C^{\mathrm{k,f}}$ $r^{\mathrm{k,f}}$ depend on data samples from the children view.

The goal of our TXQEDA+WCCN is to achieve better separation of the learned knowledge-based information. TXQEDA+WCCN replaces the two covariance matrices $\Sigma_{\mathrm{E}}^{\mathrm{k}}$ and $\Sigma_{\mathrm{I}}^{\mathrm{k}}$ of the equation (3) with $exp(\Sigma_{\mathrm{E}}^{\mathrm{k}})$ and $exp(\Sigma_{\mathrm{I}}^{\mathrm{k}})$, respectively. Thus, the objective function of TXQEDA+WCCN becomes:

$$\begin{aligned} J(W_{\mathrm{k}}) &= \frac{\mathrm{Tr}\left(W_{\mathrm{k}}^T exp(\Sigma_{\mathrm{E}}^{\mathrm{k}}) W_{\mathrm{k}}\right)}{\mathrm{Tr}\left(W_{\mathrm{k}}^T exp(\Sigma_{\mathrm{I}}^{\mathrm{k}}) W_{\mathrm{k}}\right)} \\ &= \frac{\mathrm{Tr}\left(W_{\mathrm{k}}^T \left((\mathrm{V}_{\mathrm{E}}^{\mathrm{k}})^T exp(\Lambda_{\mathrm{E}}^{\mathrm{k}})(\mathrm{V}_{\mathrm{E}}^{\mathrm{k}})\right) W_{\mathrm{k}}\right)}{\mathrm{Tr}\left(W_{\mathrm{k}}^T \left((\mathrm{V}_{\mathrm{I}}^{\mathrm{k}})^T exp(\Lambda_{\mathrm{I}}^{\mathrm{k}})(\mathrm{V}_{\mathrm{I}}^{\mathrm{k}})\right) W_{\mathrm{k}}\right)} \end{aligned} \tag{6}$$

Where :

$\Sigma_{\mathrm{E}}^{\mathrm{k}} = (\mathrm{V}_{\mathrm{E}}^{\mathrm{k}})^T \Lambda_{\mathrm{E}}^{\mathrm{k}} (\mathrm{V}_{\mathrm{E}}^{\mathrm{k}})$

$\Sigma_{\mathrm{I}}^{\mathrm{k}} = (\mathrm{V}_{\mathrm{I}}^{\mathrm{k}})^T \Lambda_{\mathrm{I}}^{\mathrm{k}} (\mathrm{V}_{\mathrm{I}}^{\mathrm{k}})$

$V_{\mathrm{E}}^{\mathrm{k}} = ((\mathbf{v}_{\mathrm{E}})_1^{\mathrm{k}}, (\mathbf{v}_{\mathrm{E}})_2^{\mathrm{k}}, \ldots, (\mathbf{v}_{\mathrm{E}})_{\mathrm{I_k}}^{\mathrm{k}})$ is the eigenvector matrix of $\Sigma_{\mathrm{E}}^{\mathrm{k}}$.

$\Lambda_{\mathrm{E}}^{\mathrm{k}}$=$diag((\lambda_{\mathrm{E}})_1^{\mathrm{k}}, (\lambda_{\mathrm{E}})_2^{\mathrm{k}}, \ldots, (\lambda_{\mathrm{E}})_{\mathrm{I_k}}^{\mathrm{k}})$ represent the corresponding eigenvalues.

$V_{\mathrm{I}}^{\mathrm{k}} = ((\mathbf{v}_{\mathrm{I}})_1^{\mathrm{k}}, (\mathbf{v}_{\mathrm{I}})_2^{\mathrm{k}}, \ldots, (\mathbf{v}_{\mathrm{I}})_{\mathrm{I_k}}^{\mathrm{k}})$ is is the eigenvector matrix of $\Sigma_{\mathrm{I}}^{\mathrm{k}}$.

$\Lambda_{\mathrm{I}}^{\mathrm{k}}$=$diag((\lambda_{\mathrm{I}})_1^{\mathrm{k}}, (\lambda_{\mathrm{I}})_2^{\mathrm{k}}, \ldots, (\lambda_{\mathrm{I}})_{\mathrm{I_k}}^{\mathrm{k}})$ represent the corresponding eigenvalues.

Therefore, there is a greater difference in the distribution scale within the class and between classes, where $exp(\Sigma_{\mathrm{E}}^{\mathrm{k}})/exp(\Sigma_{\mathrm{I}}^{\mathrm{k}})$ is larger than $\Sigma_{\mathrm{E}}^{\mathrm{k}}/\Sigma_{\mathrm{I}}^{\mathrm{k}}$. This leads to a better separation. The exponential transformation maps move the covariance matrices $\Sigma_{\mathrm{E}}^{\mathrm{k}}$ and $\Sigma_{\mathrm{I}}^{\mathrm{k}}$ to another nonlinear space:

$$\begin{aligned} &\Omega : \Re^{\mathrm{g}\times\mathrm{g}} \rightarrow \Re^{\mathrm{g}\times\mathrm{g}} \\ &\Sigma_{\mathrm{I}}^{\mathrm{k}} \rightarrow \Omega(\Sigma_{\mathrm{I}}^{\mathrm{k}}) = exp(\Sigma_{\mathrm{I}}^{\mathrm{k}}) \\ &\Sigma_{\mathrm{E}}^{\mathrm{k}} \rightarrow \Omega(\Sigma_{\mathrm{E}}^{\mathrm{k}}) = exp(\Sigma_{\mathrm{E}}^{\mathrm{k}}) \end{aligned} \tag{7}$$

Thus, TXQEDA yields to better performance compared to the TXQDA method when dealing with the nonlinearity problem. The objective function for distinctive TXQDA [11] criterion in equation (3) is to augment the same-class distance ($distance_{\mathrm{E}}$) and reduce the different-class distance ($distance_{\mathrm{I}}$) for all

tensor k^{th} mode. ($distance_{\mathrm{I}}$) and ($distance_{\mathrm{E}}$) computed by trace of covariance matrices $\Sigma_{\mathrm{E}}^{\mathrm{k}}$ and $\Sigma_{\mathrm{I}}^{\mathrm{k}}$:

$$\begin{aligned} distance_{\mathrm{E}} &= trace(\Sigma_{\mathrm{E}}^{\mathrm{k}}) \\ &= (\lambda_{\mathrm{E}})_{1}^{\mathrm{k}} + (\lambda_{\mathrm{E}})_{2}^{\mathrm{k}} + \ldots + (\lambda_{\mathrm{E}})_{\mathrm{I_k}}^{\mathrm{k}} \end{aligned}$$

and

$$\begin{aligned} distance_{\mathrm{I}} &= trace(\Sigma_{\mathrm{I}}^{\mathrm{k}}) \\ &= (\lambda_{\mathrm{I}})_{1}^{\mathrm{k}} + (\lambda_{\mathrm{I}})_{2}^{\mathrm{k}} + \ldots + (\lambda_{\mathrm{I}})_{\mathrm{I_k}}^{\mathrm{k}} \end{aligned}$$

By applying exponential function $distance_{\mathrm{E}}$ and $distance_{\mathrm{I}}$ can be calculated as follows:

$$\begin{aligned} \acute{distance}_{\mathrm{E}} &= trace(exp(\Sigma_{\mathrm{E}}^{\mathrm{k}})) \\ &= exp(\lambda_{\mathrm{E}})_{1}^{\mathrm{k}} + exp(\lambda_{\mathrm{E}})_{2}^{\mathrm{k}} + \ldots + exp(\lambda_{\mathrm{E}})_{\mathrm{I_k}}^{\mathrm{k}} \end{aligned}$$

and

$$\begin{aligned} \acute{distance}_{\mathrm{I}} &= trace(exp(\Sigma_{\mathrm{I}}^{\mathrm{k}})) \\ &= exp(\lambda_{\mathrm{I}})_{1}^{\mathrm{k}} + exp(\lambda_{\mathrm{I}})_{2}^{\mathrm{k}} + \ldots + exp(\lambda_{\mathrm{I}})_{\mathrm{I_k}}^{\mathrm{k}} \end{aligned}$$

However, TXQEDA uses the exponential process to discriminate the classes. More specifically, TXQEDA obtains the discriminant eigenvalues $exp(\Lambda_{\mathrm{E}}^{\mathrm{k}})$ and $exp(\Lambda_{\mathrm{I}}^{\mathrm{k}})$ in $\Sigma_{\mathrm{E}}^{\mathrm{k}}$ and $\Sigma_{\mathrm{I}}^{\mathrm{k}}$ such that they maximize the ratio $\frac{\lambda_{\mathrm{E_i}}}{\lambda_{\mathrm{I_i}}}$ to $\frac{exp(\lambda_{\mathrm{E_i}})}{exp(\lambda_{\mathrm{I_i}})}$ where $\frac{\lambda_{\mathrm{E_i}}}{\lambda_{\mathrm{I_i}}} << \frac{exp(\lambda_{\mathrm{E_i}})}{exp(\lambda_{\mathrm{I_i}})}$. Figure 1 illustrates the exponential application on different eigenvalues. The cost function $J(W_{\mathrm{k}})$ in (6) can be rewritten by the generalized eigenvalue decomposition:

$$exp(\Sigma_{\mathrm{E}}^{\mathrm{k}})W_{\mathrm{k}} = \Lambda_{\mathrm{k}} exp(\Sigma_{\mathrm{I}}^{\mathrm{k}})W_{\mathrm{k}} \tag{8}$$

Where, W_{k} is the eigenvectors matrix and Λ_{k} is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_{\mathrm{i}}^{\mathrm{k}}$.

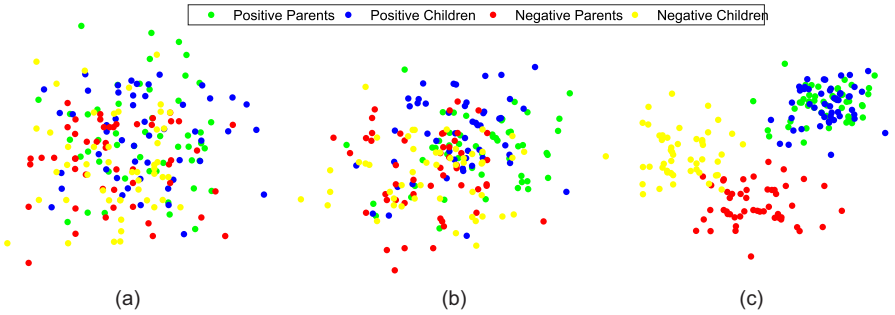


Figure 6: Illustration of margin maximization between samples belonging to positive and negative pairs. (a) The original face images with/without kinship relations in the primary Mv-VGG space. (b) The face images after projection through the basic TXQDA into the new high-dimensional space. (c) The face images after projection by the proposed method (TXQEDA+WCCN) into the new high-dimensional space.

To illustrate the mechanism of how TXQEDA affects the performance of kin verification, we draw the nodes of different elements (images) with different colors in 2-D space. Each point represents a 2-D representation of the corresponding sample, with the green, blue, red, and yellow points denoting the positive parents, positive children, negative parents, and negative children, respectively. Figure 6 shows the results of visualizing our algorithm and baselines. In Figure 6-(a), we can see that the original face images spread in the Mv-VGG space with large overlap. In Figure 6-(c), we can see that TXQEDA+WCCN distinguishes the nodes with same/different labels more clearly than the baseline algorithms (TXQDA) in Figure 6-(b). The result of TXQDA (Figure 6-(b)) is not satisfactory because all the negative nodes are intermixed and the positive ones are spaced. In TXQEDA+WCCN (Figure 6-(c)), most of the negative nodes are separated and the positive ones are contiguous. However, the positive pairs are projected as close as possible and the negative ones are pushed as far away as possible. In this way, the distance between the negative pairs is maximized and the distance between positive pairs is minimized. It is observed that the visualization of the separation result by TXQEDA is better compared with the baselines.

4.2. Within Class Covariance Normalization

WCCN [31] metric is one of the most preferred metrics to decrease the effect of the within class variations in Speaker Recognition literature. However, it uses the minimization of within class variability for feature learning, which reduces the expected variances between training features of the same class. The WCCN metric has highlighted the benefit of mapping the feature vectors of the TXQEDA approach to a new subspace. Therefore, it is of great importance for us to obtain an efficient version of TXQEDA. We show that TXQEDA+WCCN can efficiently learn multi-view deep features to handle the complex variability in the same classes.

$$
\begin{aligned}
{}_{\mathrm{n_I}}G^{\mathrm{k}} &= \sum_{\mathrm{f}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}}\mathrm{I_o}} {}_{\mathrm{n_I}}G^{\mathrm{k,f}} \\
{}_{\mathrm{n_I}}G^{\mathrm{k,f}} &= (W^{\mathrm{k}})^T\tilde{P}^{\mathrm{k,f}}(\tilde{P}^{\mathrm{k,f}})^T + (W^{\mathrm{k}})^T\tilde{C}^{\mathrm{k,f}}(\tilde{C}^{\mathrm{k,f}})^T - (W^{\mathrm{k}})^T S^{\mathrm{k,f}}(R^{\mathrm{k,f}})^T - (W^{\mathrm{k}})^T R^{\mathrm{k,f}}(S^{\mathrm{k,f}})^T
\end{aligned}
\tag{9}
$$

W^{k} represents the projection matrix estimated in (8), the WCCN matrix Q can be calculated and expressed by Cholesky decomposition [32] as:

$$
(G^{\mathrm{k}})^{-1} = Q^{\mathrm{k}}(Q^{\mathrm{k}})^T \tag{10}
$$

Hence, we have the following new projection matrix W^{k} :

$$
W^{\mathrm{k}} = (Q^{\mathrm{k}})^T W^{\mathrm{k}} \tag{11}
$$

TXQEDA uses an iterative optimization process to estimate the projection matrices in equation (6), making the proposed distance metric more stable and efficient during the classification process. TXQEDA learns a subspace $W_1, W_2, \ldots, W_{\mathrm{k}-1}, W_{\mathrm{k}+1}, \ldots W_{\mathrm{N}}$ that supposed known and initialized to identity initially and W_{k} is estimated. Set: $\boldsymbol{U} = \boldsymbol{A} \times_1 W_1 \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1} \ldots \times_{\mathrm{N}} W_{\mathrm{N}}$ and $\boldsymbol{V} = \boldsymbol{B} \times_1 W_1 \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1} \ldots \times_{\mathrm{N}} W_{\mathrm{N}}$, the TXQEDA algorithm has to be formulated in terms of iterative optimization the discriminative subspace $\boldsymbol{A}$ and $\boldsymbol{B}$ into $\boldsymbol{U}$ and $\boldsymbol{V}$ in equation (4) and equation (5), respectively. The optimization process terminates if any of the following

conditions are met: (a) The number of iterations reaches a predefined maximum; or b) The difference in the estimated projection between two successive iterations is smaller than a threshold, $\left\|W_{\mathrm{k}}^{\mathrm{iter}} - W_{\mathrm{k}}^{\mathrm{iter}-1}\right\| < \mathrm{I_k I_k}\epsilon$, where iter refers to the current iteration number and $\mathrm{I_k}$ is the k^{th} mode dimension of $W_{\mathrm{k}}^{\mathrm{iter}}$.

The TXQEDA+WCCN parameters $\mathrm{i}_1 \times \mathrm{i}_2 \times \ldots \times \mathrm{i}_{\mathrm{N}}$ (final dimensions) are solved by applying the energy percentage ($\mathrm{Energy}^{\mathrm{k}}$) to the eigenvalues for each tensor mode:

$$\mathrm{Energy}^{\mathrm{k}} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{i_k}} \lambda_{\mathrm{j}}^{\mathrm{k}}}{\sum_{\mathrm{j}=1}^{\mathrm{I_k}} \lambda_{\mathrm{j}}^{\mathrm{k}}} \times 100 \tag{12}$$

with $\lambda_1^{\mathrm{k}} > \lambda_2^{\mathrm{k}} > \ldots > \lambda_{\mathrm{I_k}}^{\mathrm{k}}$.

The entire procedure of the proposed TXQEDA+WCCN scheme is described in Algorithm 1.

Algorithm 1 TXQEDA+WCCN

Input:

- The tensor $\boldsymbol{A} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{N} \times \mathrm{n}}$ contains parents samples (the first view).
- The tensor $\boldsymbol{B} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{N} \times \mathrm{m}}$ contains children samples samples (the second view).
- The positive labels ($\mathrm{labels}_\mathrm{W}$) for extracting the match tensor pairs.
- The maximal number of iterations $\mathrm{Iteration}_{\max}$.
- The final lower dimensions: $\mathrm{i}_1 \times \mathrm{i}_2 \times \cdots \times \mathrm{i}_\mathrm{N}$.

Output:

- The projection matrices of each mode $W_\mathrm{k} = W_\mathrm{k}^{\mathrm{iter}} \in \Re^{\mathrm{I_k} \times \mathrm{i_k}}, \mathrm{k} = 1, \cdots, \mathrm{N}$.

Algorithm:

1. **Initialization:** $W_1^0 = I_{\mathrm{I}_1}, W_2^0 = I_{\mathrm{I}_2}, \cdots, W_\mathrm{N}^0 = I_{\mathrm{I}_\mathrm{N}}$.
2. **For** iter : 1 to $\mathrm{Iteration}_{\max}$
   (a) **For** k=1 to N
   - $\boldsymbol{U} = \boldsymbol{A} \times_1 W_1^{\mathrm{iter}-1} \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1}^{\mathrm{iter}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1}^{\mathrm{iter}-1} \ldots \times_\mathrm{N} W_\mathrm{N}^{\mathrm{iter}-1}$.
   - $U^\mathrm{k} \Leftarrow_\mathrm{k} \boldsymbol{U}$.
   - $\boldsymbol{V} = \boldsymbol{B} \times_1 W_1^{\mathrm{iter}-1} \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1}^{\mathrm{iter}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1}^{\mathrm{iter}-1} \ldots \times_\mathrm{N} W_\mathrm{N}^{\mathrm{iter}-1}$.
   - $V^\mathrm{k} \Leftarrow_\mathrm{k} \boldsymbol{V}$.
   - $\mathrm{n_I}\Sigma_\mathrm{I}^\mathrm{k} = \sum_{\mathrm{f}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_I}\Sigma_\mathrm{I}^{\mathrm{k,f}}$.
     $\mathrm{n_I}\Sigma_\mathrm{I}^{\mathrm{k,f}} = \tilde{P}^{\mathrm{k,f}}(\tilde{P}^{\mathrm{k,f}})^T + \tilde{C}^{\mathrm{k,f}}(\tilde{C}^{\mathrm{k,f}})^T - S^{\mathrm{k,f}}(R^{\mathrm{k,f}})^T - R^{\mathrm{k,f}}(S^{\mathrm{k,f}})^T$.
   - $\mathrm{n_E}\Sigma_\mathrm{E}^\mathrm{k} = \sum_{\mathrm{f}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_E}\Sigma_\mathrm{E}^{\mathrm{k,f}}$.
     $\mathrm{n_E}\Sigma_\mathrm{E}^{\mathrm{k,f}} = \mathrm{m}P^{\mathrm{k,f}}(P^{\mathrm{k,f}})^T + \mathrm{n}C^{\mathrm{k,f}}(C^{\mathrm{k,f}})^T - \mathbf{s}^{\mathrm{k,f}}(\mathbf{r}^{\mathrm{k,f}})^T - \mathbf{r}^{\mathrm{k,f}}(\mathbf{s}^{\mathrm{k,f}})^T - \mathrm{n_I}\Sigma_\mathrm{I}^{\mathrm{k,f}}$.
   - **Moving the covariance matrices $\Sigma_\mathrm{E}^\mathrm{k}$ and $\Sigma_\mathrm{I}^\mathrm{k}$ into another nonlinear space by:**
     $\Omega : \Re^{\mathrm{g}\times\mathrm{g}} \rightarrow \Re^{\mathrm{g}\times\mathrm{g}}$
     $\Sigma_\mathrm{I}^\mathrm{k} \rightarrow \Omega(\Sigma_\mathrm{I}^\mathrm{k}) = exp(\Sigma_\mathrm{I}^\mathrm{k})$
     $\Sigma_\mathrm{E}^\mathrm{k} \rightarrow \Omega(\Sigma_\mathrm{E}^\mathrm{k}) = exp(\Sigma_\mathrm{E}^\mathrm{k})$
   - **Estimate $W_\mathrm{k}^{\mathrm{iter}}$ by solving the following eigenvalue decomposition problem :**
     $exp(\Sigma_E^\mathrm{k}) W_\mathrm{k}^{\mathrm{iter}} = \Lambda_\mathrm{k} exp(\Sigma_\mathrm{I}^\mathrm{k}) W_\mathrm{k}^{\mathrm{iter}}$.
   - $\mathrm{n_I}G^\mathrm{k} = \sum_{\mathrm{f}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_I}G^{\mathrm{k,f}}$.
     $\mathrm{n_I}G^{\mathrm{k,f}} = (W^\mathrm{k})\tilde{P}^{\mathrm{k,f}}(\tilde{P}^{\mathrm{k,f}})^T + (W^\mathrm{k})\tilde{C}^{\mathrm{k,f}}(\tilde{C}^{\mathrm{k,f}})^T - (W^\mathrm{k})S^{\mathrm{k,f}}(R^{\mathrm{k,f}})^T - (W^\mathrm{k})R^{\mathrm{k,f}}(S^{\mathrm{k,f}})^T$.
   - **Compute WCCN projection matrix** Q^k: $(G^\mathrm{k})^{-1} = Q^\mathrm{k}(Q^\mathrm{k})^T$
   - **Compute the new projection matrix** $W_\mathrm{k}^{\mathrm{iter}} = (Q^\mathrm{k})^T W_\mathrm{k}$

   (b) **If** iter > 2 and $\left\| W_\mathrm{k}^{\mathrm{iter}} - W_\mathrm{k}^{\mathrm{iter}-1} \right\| < \mathrm{I_k I_k}\epsilon, \mathrm{k} = 1, \cdots, \mathrm{N}$, break;
3. Compute the final lower dimensions $\mathrm{i_k}$ by: $\mathrm{Energy}^\mathrm{k} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{i_k}} \lambda_\mathrm{j}^\mathrm{k}}{\sum_{\mathrm{j}=1}^{\mathrm{I_k}} \lambda_\mathrm{j}^\mathrm{k}} \times 100$, where ($\lambda_1^\mathrm{k} > \lambda_2^\mathrm{k} > \ldots > \lambda_{\mathrm{I_k}}{}^\mathrm{k}$).
4. Sort the $\mathrm{i_N}$ eigenvectors $W_\mathrm{k}^{\mathrm{iter}} \in \Re^{\mathrm{I_k}\times\mathrm{i_k}}$ according to Λ_k in decreasing order, $\mathrm{k} = 1, \cdots, \mathrm{N}$.

5. Performance evaluation

The proposed kin verification approach has been experimentally verified by applying it to publicly available kin verification datasets(KinFaceW-I, KinFaceW-II [1], TSKinFace [33], and Cornell KinFace database [34]). Images in these databases were captured and recorded under uncontrolled conditions for gestures, demographic attributes, background, lighting, partial occlusion, and facial expressions. The KinFaceW-I dataset contains 4 kinship relations. There are 156 (F-S), 134 (F-D), 116 (M-S) and 127 (M-D) kinship pairs. The KinFaceW-II dataset contains 4 kinship relations (F-S), (F-D), (M-S) and (M-D), 250 pairs of kinship images for each kinship relation. We also used the TSKinFace database, which contains two types of kinship relations between three individuals: Father-Mother-Daughter (FM-D) with 502 relationships and Father-Mother-Son (FM-S) with 513 relationships. To make a fair comparison, we reorganized the TSKinFace database by splitting the Father-mother-daughter group into two groups father-daughter (F-D) with 502 relationships, Mother-Daughter (M-D) with 502 kinship relations, and the Father-Mother-Son group into two groups Father-Son (F-S) with 513 kinship relations, Mother-Son (M-S) with 513 kinship relations, in total the TSKinFace database contains 2052 face images. The Cornell KinFace database contains 150 pairs (300 images) with different demographic attributes. The distribution of kinship pairs is as follows: 40% (F-S), 22% (F-D), 13% (M-S), and 26% (M-D).

5.1. Parameter Settings

We followed the framework and testing protocol in [1] and used a five-fold cross-validation testing scheme to evaluate the performance of our approach. For the four kinship databases KinFaceW-I, KinFaceW-II [1], TSKinFace [33], and the Cornell KinFace database [34], we cropped the images to face size 64×64 pixels, then we resized the input images to the required 224×224 dimensions to use as input for deep feature extraction. The TXQEDA+WCCN subspaces are set to 96% energy of the eigenvalues, preserving the eigenval-

ues that provide up to 96% of the information. For example, in KinFaceW-II, and for the F-S relation, the $\boldsymbol{A}$, $\boldsymbol{B}$ are $\in \Re^{4096\times 8\times 200}$ and their projection $1000 \times 8 \times 200$, and also provide the dimensions of the two projections W_1 with dimensions 8×8 and W_2 with dimensions 4096×1000.

5.2. Experimental Results

We tested our TXQEDA+WCCN approach on the various relations. Table 1 (KinFaceW-I), Table 2 (KinFaceW-II), Table 3 (TSKinFace), Table 4 (Cornell KinFace) show the performance of the proposed TXQEDA+WCCN approach with different deep feature descriptors. The results of TXQEDA+WCCN performance compared to the state-of-the-art methods are shown in Table 5 (KinFaceW-I), Table 6 (KinFaceW-II), Table 7 (TSKinFace), and Table 8 (Cornell KinFace). Experiments were also performed on the baseline approaches with the same settings, which are included in Table 9 and 10. We note that:

Table 1: The average verification accuracy (%) of the proposed method using several deep features on KinFaceW-I database.

Method	Features	F-S	F-D	M-S	M-D	Mean
TXQDA	VGG-F	80.07	82.42	78.48	78.79	79.94
TXQDA	VGG-M	77.36	80.00	74.85	76.67	77.22
TXQDA	VGG-S	81.30	84.85	84.47	87.78	84.60
TXQDA	VGG-Face	74.84	71.82	76.67	76.97	75.07
TXQDA	Mv-VGG	75.77	73.33	78.48	77.88	76.36
TXQEDA+WCCN	VGG-F	82.93	70.30	81.82	79.39	78.61
TXQEDA+WCCN	VGG-M	82.00	76.67	80.00	79.09	79.44
TXQEDA+WCCN	VGG-S	83.20	79.39	81.82	84.24	82.16
TXQEDA+WCCN	VGG-Face	73.99	68.48	75.15	73.64	72.81
TXQEDA+WCCN	**Mv-VGG**	**91.00**	**87.78**	**92.32**	**93.35**	**91.11**

- From the analysis of the results associated with the four databases Table 1 (KinFaceW-I), Table 2 (KinFaceW-II), Table 3 (TSKinFace) and , Table 4 (Cornell KinFace) using the Mv-VGG description, our proposed TXQEDA+WCCN method outperforms the TXQDA schemes for most of the image features. The results reported in Tables 1, 2, 3, and 4 show that

Table 2: The average verification accuracy (%) of the proposed method using several deep features on KinFaceW-II database.

Method	Features	F-S	F-D	M-S	M-D	Mean
TXQDA	VGG-F	78.60	79.80	78.40	77.60	78.60
TXQDA	VGG-M	76.40	77.80	76.00	75.60	76.45
TXQDA	VGG-S	83.40	81.20	82.40	81.20	82.05
TXQDA	VGG-Face	76.00	71.80	75.20	76.40	74.85
TXQDA	Mv-VGG	78.40	72.20	76.40	76.40	75.95
TXQEDA+WCCN	VGG-F	82.40	76.80	82.20	81.40	80.70
TXQEDA+WCCN	VGG-M	82.60	81.00	82.00	65.20	77.70
TXQEDA+WCCN	VGG-S	84.00	79.40	82.20	84.00	82.40
TXQEDA+WCCN	VGG-Face	79.20	73.40	77.60	79.00	77.30
TXQEDA+WCCN	**Mv-VGG**	**89.80**	**90.60**	**87.60**	**93.20**	**90.30**

Table 3: The average verification accuracy (%) of the proposed method using several deep features on TSKinFace database.

Method	Features	F-S	F-D	M-S	M-D	Mean	FM-S	FM-D
TXQDA	VGG-F	70.10	69.41	70.97	71.94	70.60	77.18	72.72
TXQDA	VGG-M	73.59	72.38	71.46	73.98	72.85	79.71	74.83
TXQDA	VGG-S	67.48	66.63	67.38	68.25	72.43	73.30	69.81
TXQDA	VGG-Face	74.27	74.06	78.06	79.61	76.50	83.01	79.42
TXQDA	Mv-VGG	87.08	84.45	88.25	88.69	87.11	94.37	93.85
TXQEDA+WCCN	VGG-F	68.06	66.83	67.77	69.03	67.92	72.82	69.13
TXQEDA+WCCN	VGG-M	81.84	79.70	82.52	8029	81.08	89.90	84.08
TXQEDA+WCCN	VGG-S	64.27	63.66	65.53	64.66	64.53	66.5	69.42
TXQEDA+WCCN	VGG-Face	76.31	70.29	74.66	76.31	74.39	76.99	76.80
TXQEDA+WCCN	**Mv-VGG**	**89.42**	**89.31**	**90.87**	**93.15**	**90.68**	**95.34**	**96.53**

Table 4: The average verification accuracy (%) of the proposed method using several deep features on Cornell KinFace database.

Method	Features	Accuracy
TXQDA	VGG-F	86.39
TXQDA	VGG-M	88.88
TXQDA	VGG-S	88.04
TXQDA	VGG-Face	88.87
TXQDA	Mv-VGG	92.70
TXQEDA+WCCN	VGG-F	88.44
TXQEDA+WCCN	VGG-M	88.88
TXQEDA+WCCN	VGG-S	88.37
TXQEDA+WCCN	VGG-Face	89.22
TXQEDA+WCCN	**Mv-VGG**	**93.77**

Table 5: Comparison of the proposed approach with other kinship verification approaches on KinFaceW-I database.

Method	F-S	F-D	M-S	M-D	Mean
MNRML [1]	72.50	66.50	66.20	72.00	69.90
DMML [2]	74.50	69.50	69.50	75.50	72.25
MPDFL [35]	73.50	67.50	66.10	73.10	70.10
DDMML [36]	86.40	79.10	81.40	87.00	83.50
SIEDA [37]	-	-	-	-	80.00
PML-COV-S [38]	91.00	84.30	87.10	90.20	88.20
KML [39]	83.80	81.00	81.20	85.00	82.80
MNRML+SVM [4]	85.90	79.85	86.20	86.62	84.55
KVRL+fcDBN [40]	98.10	96.30	90.50	98.40	96.10
Method in [41]	-	-	-	-	96.90
AdvKin [42]	75.70	78.30	77.60	83.10	78.70
Method in [43]	-	-	-	-	80.50
WGEML [7]	78.50	73.90	80.60	81.90	78.70
NESN-KVN [44]	76.50	77.00	85.20	75.80	78.60
Method in [45]	85.80	87.50	88.10	80.90	85.60
MSIDA+WCCN [8]	85.98	85.93	90.05	88.62	87.65
SSADL-HF [9]	87.10	79.50	81.90	83.90	83.10
TXQDA (Mv-VGG)	75.77	73.33	78.48	77.88	76.36
TXQEDA+WCCN (Mv-VGG)	**91.00**	**87.78**	**92.32**	**93.35**	**91.11**

using Mv-VGG with our method improves the accuracy of 14%, 3.5% and 1.07% for bi-subject kinship verification on KinFaceW-I, KinFaceW-II and TSKinFace databases and Cornell KinFace database, respectively. For a general observation, TXQEDA+WCCN was trained with a single feature representation to learn kinship verification, with the goal of finding the image representation that leads to the highest accuracy. Tables 1, 2, 3 and 4 show the results of these tests. Accordingly, the TXQEDA+WCCN approach performs better than the TXQDA approach on all features, both in verifying single or multi-view kin verification.

- The proposed TXQEDA+WCCN method outperforms other shallow approaches [1, 2, 11, 5, 37, 38] diffused with handcrafted features and metric learning models, as shown in Table 5 (KinFaceW-I), Table 6 (KinFaceW-II), Table 3 (TSKinFace) and Table 4 (Cornell KinFace). Note that some of these methods, such as [1, 2, 11, 5] , used features fusion to describe

Table 6: Comparison of the proposed approach with other kinship verification approaches on KinFaceW-II database.

Method	F-S	F-D	M-S	M-D	Mean
MNRML [1]	76.90	74.30	77.40	77.60	76.50
DMML [2]	78.50	76.50	78.50	79.50	78.25
MPDFL [35]	77.30	74.70	77.80	78.00	77.00
DDMML [36]	87.40	83.80	83.20	83.00	84.30
SIEDA [37]	-	-	-	-	87.60
PML-COV-S [38]	88.60	85.80	87.20	91.00	88.20
MNRML+SVM [4]	87.20	82.60	88.40	89.40	86.90
SILD+WCCN/LR [5]	88.40	84.20	85.80	86.40	86.20
KML [39]	87.40	83.60	86.20	85.60	85.70
KVRL+fcDBN [40]	96.80	94.00	97.20	96.80	96.20
Method in [41]	-	-	-	-	97.10
TXQDA [11]	90.20	86.40	85.60	86.40	87.15
AdvKin [42]	88.40	85.80	88.00	89.80	88.00
Method in [43]	-	-	-	-	82.30
WGEML [7]	88.60	77.40	83.40	81.60	82.80
NESN-KVN [44]	86.70	88.70	91.60	89.10	89.00
Method in [45]	90.40	86.60	91.00	87.20	88.80
MSIDA+WCCN [8]	89.40	82.80	87.80	88.00	87.00
SSADL-HF [9]	88.60	80.00	81.00	80.00	82.40
TXQDA (Mv-VGG)	78.40	72.20	76.80	76.40	75.95
TXQEDA+WCCN (Mv-VGG)	**89.80**	**90.60**	**87.60**	**93.20**	**90.30**

Table 7: Comparison of the proposed approach with other kinship verification approaches on TSKinFace database.

Method	F-S	F-D	M-S	M-D	Mean	FM-S	FM-D
RSBM [33]	83.00	80.50	82.80	81.10	81.85	86.40	84.40
DDMML [36]	86.60	82.50	83.20	84.30	84.15	88.50	87.10
MKSM [46]	84.80	83.20	85.19	84.90	84.52	-	-
SILD+WCCN/LR [5]	89.08	87.05	88.59	89.63	88.59	90.94	91.23
TXQDA [11]	89.32	90.69	90.29	90.97	90.32	94.85	95.63
WGEML [7]	90.30	89.80	91.40	90.40	90.47	93.50	93.00
TXQDA (Mv-VGG)	87.08	84.45	88.25	88.69	87.11	94.37	93.85
TXQEDA+WCCN (Mv-VGG)	**89.42**	**89.31**	**90.87**	**93.15**	**90.68**	**95.34**	**96.53**

the face images.

- The proposed framework also outperforms Deep Learning based face verification for kinship verification, such as MNRML+SVM [5], AdvKin[42], KML [39], DDMML [36], WGEML [7]. Unlike these, the method in [41]

Table 8: Comparison of the proposed approach with other kinship verification approaches on Cornell KinFace database.

Method	Accuracy
DMML [2]	73.75
MPDFL [35]	71.90
SIEDA [37]	81.50
KVRL+fcDBN [40]	89.75
AdvKin [42]	81.40
KML [39]	81.40
TXQDA [11]	93.04
Method in [41]	**94.40**
SSADL-HF [9]	70.30
TXQDA (Mv-VGG)	92.70
TXQEDA+WCCN (Mv-VGG)	**93.77**

outperforms all approaches to kinship verification approaches in KinFace-I, KinFace- II, and the Cornell KinFace database. Also, the KVRL+fcDBN [40] approach outperforms ours and other CNN-based methods except for the method in [41] in KinFace-I, KinFace-II database. The reason is that the work in [40] is based on a deep learning approach (KVRL-fcDBN), the fcDBN algorithm was to learn a large number of external data (more than 600,000 external face images) and achieved 96.20% on KinFaceW-II databases, 96.10% on KinFaceW-I databases, and 89.50% on Cornell KinFace. The authors in [41] used an external dataset in a new kinship video database (KIVI) that contains over 250.000 still images extracted from 355 real kinship video pairs, and achieved 97.10% on KinFaceW-II, 96.90% on KinFaceW-I, and 94.40% on Cornell KinFace. However, in our work, we used tensor approach TXQEDA+WCCN to learn only the available data (i.e., no external data used in the training phase) and we achieved performances 90.30% on the KinFaceW-II database, 91.16% on the KinFaceW-I database, 90.68% on the TSKinFace database, and 93.77% on the Cornell KinFace database. In general, CNN-based methods provide low performance when using small datasets. The 4 VGG models (i.e., VGG-M, VGG-F, VGG-Face, and VGG-S) also used external images like the Imagenet and VGGFace datasets, but only for object features and

face features, respectively. In fact, training these CNN architectures did not use external kinship (i.e., mother-son, father-son, mother-daughter, and father-daughter, etc.) to train these models to solve kinship verification problems. While KVRL+fcDBN [40] and the method in [41], external faces are used according to policies related only to the kinship verification problem. In other words, kinship relations were specifically used in the development of the algorithm in [40, 41].

- The proposed TXQEDA+WCCN framework outperforms all the kinship methods except the methods in [41], KVRL+fcDBN [40] which are used on external large face data. This proves the effectiveness of using the robust tensor model (TXQEDA+WCCN) on kinship verification topics. Moreover, TXQEDA+WCCN improved the performance of their counterpart (i.e. TXQDA). Thus, the integration of the exponential criterion and the WCCN metric leads to stable and high performances in kinship verification.

5.3. Discussion and further experiments

- TXQEDA+WCCN shows the effectiveness of combining face and object features in tensor models. This is because combining deep features can provide complementary information for feature learning which is robust enough to represent the samples used for training TXQEDA+WCCN. The VGG-Face models [23] are trained on face images containing 2.6 million images of over 2.6K individuals to extract deep facial features manifested in deep information about identity, age, gender, ethnicity, expression, illumination, and pose. While the models trained on object images (i.e., VGG-S, VGG-F, and VGG-M) [22] with the large number of object images, these models help to extract the key hidden facial features that manifest in deep information about: contours, edges, corners, shape etc., extracted from the face parts (i.e., eyes, eyebrows, nose, mouth, etc.).

- In contrast to the TXQEDA+WCCN approach, most existing kinship veri-

fication methods use hand-crafted low-level descriptors (Shallow features), which are unlikely to be powerful enough to characterize kinship based on facial images. Also, the other kinship verification methods that use deep features simply do not take into account the problem of cross-view matching and the curse of intra-class variability.

- For the databases provided, the competing methods reached an accuracy of 66% to 93.77%, without counting the results of KVRL+fcDBN [40] and the method in [41], the proposed TXQEDA+WCCN obtained the best results (90.30% for the KinFaceW-II database, 91.16% for the KinFaceW-I database, 90.68% for the TSKinFace database, and 93.77% for the Cornell KinFace database). To make a fair comparison, we need to take into account the external data used to train kinship verification methods. The methods that used a large external face database to train the algorithm (i.e. KVRL+fcDBN [40], method in [41]) were able to increase the average verification accuracy of kinship verification. We believe that the results can be higher and exceed the results of KVRL+fcDBN [40] and the method in [41] when external data are used to train our TXQEDA+WCCN.

- All parent-child pairs in the KinFaceW-II dataset are cropped from the same source. This fact of cropping seems to simplify the classification considerably. The KinFaceW-II dataset contains a larger number of photos cropped from the same original image compared to KinFaceW-I. The authors in [47] concluded that the KinFaceW-II dataset is less challenging than the KinFaceW-I dataset. Therefore, we have greater confidence results for KinFaceW-I dataset.

- To gain more confidence in the efficacy of our system in face pruning, we also evaluated our approach using another database. In the Cornell KinFace database, most pairs were obtained from different images from different cameras and different times under uncontrolled environment. From table 8, our approach is clearly outperforming recent methods for kinship verification.

To better illustrate the high performance and effectiveness of our TXQEDA+WCCN and its TXQDA counterpart, the receiver operating characteristic (ROC) curves of the different deep descriptors are shown in Figures 7, 8, 9, and 10, respectively. 10, respectively. We observe that the proposed fusion scheme of deep features (Mv-VGG) with the TXQEDA+WCCN tensor approach provided the best performance in terms of the ROC curve for all tasks.

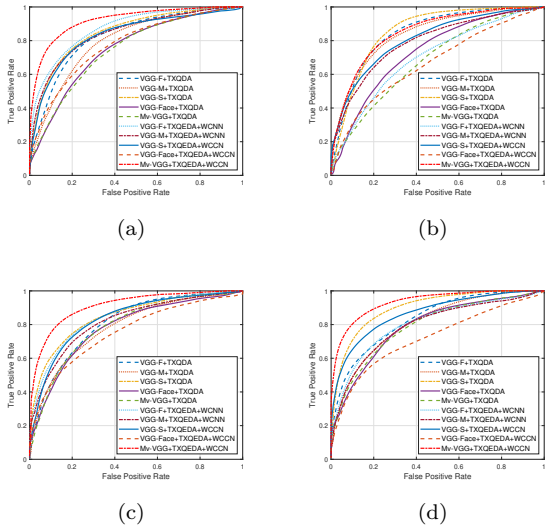


Figure 7: ROC curves of different deep features descriptors with our proposed method on KinFaceW-I database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

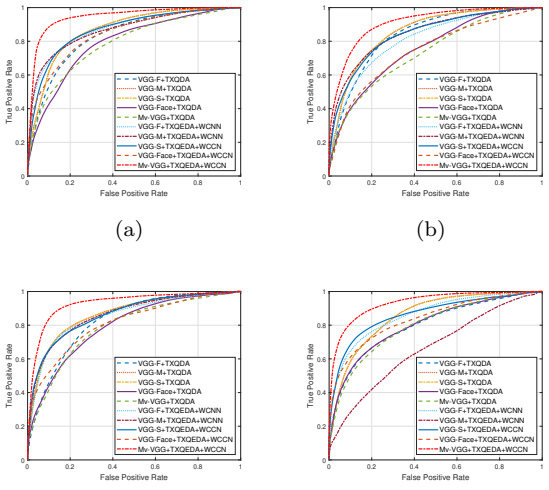


Figure 8: ROC curves of different deep features descriptors with our proposed method on KinFaceW-II database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

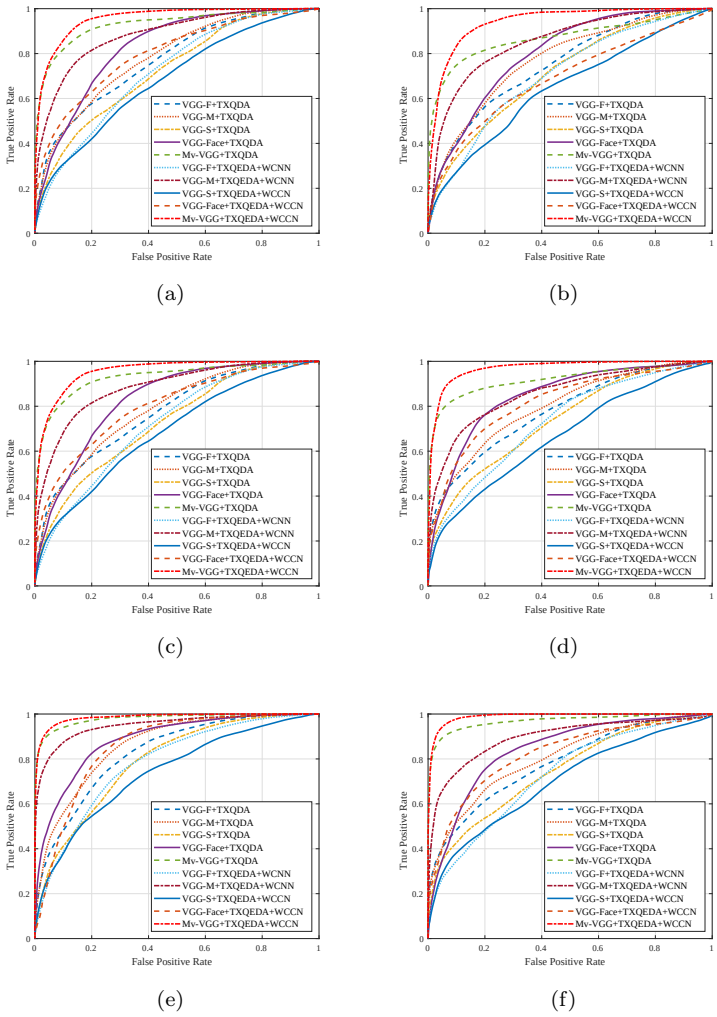
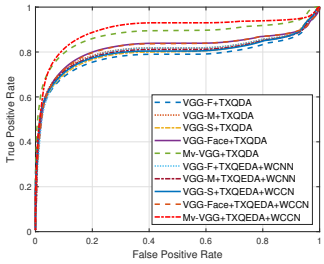


Figure 9: ROC curves of different deep features descriptors with our proposed method on TSKinFace database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, (e) FM-S set and (f) FM-D set, respectively.



(a)

Figure 10: ROC curves of different deep features descriptors with our proposed method on Cornell database.

- **Our approach vs. TXQDA method:**

  In both cases, our TXQEDA+WCCN approach and the TXQDA method, features are structured in the shape of a tensor of order 3 in the hope of extracting maximum information from the input face image. The basic TXQDA approach has been applied in a multi-view learning system and works well only for shallow features. Unlike the deep features, the shallow features do not cause the problem of intra-class variability which increases the classification error. The intra-class variability causes TXQDA to classify the pairs that are related as unrelated pairs. On the other hand, this error is significantly reduced when the proposed TXQEDA+WCCN method is used. We emphasize that a single deep descriptor generates one level of intra-class variability, while deep descriptors with multiple views generate multiple levels of intra-class variability. In this case, classification is more difficult for TXQDA as the classification error may increase. Incorporating the WCCN metric into our approach helps to solve the intra-class variability problem, thereby reducing the classification error. To further increase the inter-sample distance, we included the exponential matrix in the TXQEDA+WCCN approach. Moreover, by using the exponential matrices, the TXQEDA+WCCN method is not affected by the nonlinearity problem.

- **Our tensor-based (multi-linear) method vs. vector-based (linear) methods using multi-view features:**

  Many methods, such as [1, 2, 5, 9], use several different features by transforming them into a linear subspace to project the face images used for kinship verification. In general, applying the linear subspace transformation to features with multiple views compromises the natural structure of the data, which is neglected or lost when arranging the feature vectors. Moreover, each feature type may lose its discriminative information due to concatenation with other feature types used by linear subspace transform methods. Therefore, these methods arrange multiple data regions into a subspace (linear subspace). These methods may be limited in that they cannot explicitly take any property from the description of multiple features for each view. Unlike the linear subspace transformation, the transformation into multiple linear subspaces (i.e., our TXQEDA+WCCN method) preserves the data structure by stacking the data in a tensor representation, which provides more discriminative information. Moreover, with a high order tensor subspace, TXQEDA+WCCN can circumvent the dimensionality dilemma by using the k-mode optimization technique. This optimization technique helps to obtain a more discriminative subspace with smaller dimension than traditional linear multi-view methods.

- **Our approach vs. baseline methods:**

  To quantify the actual effect of single/multiple types of deep features, we used simple score matching separately. Simple score matching is based on estimating the kin relation between samples directly using the cosine similarity between the trait vectors of the parents and the children. The motivation for using cosine similarity is to demonstrate the true effect of single/multiple deep features. At this point, cosine similarity is used with the original data without any pre-processing steps (calculating the cosine similarity between the family members' feature vectors before any projection). As can be seen in Table 9, the cosine similarity has obtained the

lowest result for all the feature tasks. This means that we cannot assume that the actual performance improvement achieved by our framework is only due to the single/multiple deep features.

In contrast to cosine similarity, the main motivation for using the proposed scheme for the kinship verification problem is to reduce the difference between parents and children in order to decide whether or not they later come from the same family. The projection matrices of our method reduce the difference between parents and children who are really from the same family, and increase the difference between parents and children who had no kinship relationship. The goal is achieved in three ways (preprocessing before we send them to cosine similarity): (i) deep feature extraction, dimensionality reduction and margin maximization based on multi-linear representation and exponential discriminant analysis criterion; (ii) feature fusion of multiple deep features in different modes; (iii) solving the intra-class variability problem caused by one/multiple deep features. From Table 9, we can see that the effectiveness of single/multiple deep features is given only by a robust knowledge-based tensor model (TXQEDA+WCCN)

Table 9: Performance comparison (%) of simple score matching and TXQEDA+WCCN using different deep features on KinFaceW-I, KinFaceW-II, TSKinFace and Cornell datasets.

Method	Features	KinFaceW-I					KinFaceW-II					TSKinFace					Cornell
		F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean	Mean
Simple score matching	VGG-F	74,89	81.00	75,83	78,32	77,51	72,60	74.00	72,40	72,60	72,90	70,47	68,43	69,88	71,21	69,99	72,42
	VGG-M	71,39	77,25	72,77	75,95	74,34	70.00	72,20	70,60	69,80	70,65	71,44	68,82	70,76	72,41	70,85	72,78
	VGG-S	75,90	82,08	78,86	81,51	79,58	76,80	77,80	77,20	76.00	76,95	72,32	70,71	73,09	74,60	72,68	73,03
	VGG-Face	70,90	70,09	76,68	79,08	74,18	71,80	68,60	73,40	73.00	71,70	68,22	68,02	68,42	73,10	69,44	71,80
	Mv-VGG	71,57	70,83	77,12	79,88	74,85	72,80	69.00	73,80	73,80	72,35	68,42	67,92	68,72	73,10	69,54	71,77
TXQEDA+WCCN	VGG-F	82,93	70,30	81,82	79,39	78,61	82,40	76,80	82,20	81,40	80,70	68,06	66,83	67,77	69,03	67,92	88,44
	VGG-M	82.00	76,67	80.00	79,09	79,44	82,60	81.00	82.00	65,20	77,70	81,84	79,70	82,52	80.29	81,08	88,88
	VGG-S	83,20	79,39	81,82	84,24	82,16	84.00	79,40	82,20	840	82,40	64,27	63,66	65,53	64,66	64,53	88,37
	VGG-Face	73,99	68,48	75,15	73,64	72,81	79,20	73,40	77,60	79.00	77,30	76,31	70,29	74,66	76,31	74,39	89,22
	Mv-VGG	**91.00**	**87,78**	**92,32**	**93,35**	**91,11**	**89,80**	**90,60**	**87,60**	**93,20**	**90,30**	**89,42**	**89,31**	**90,87**	**93,15**	**90,68**	**93,77**

Table 10: Performance comparison (%) of baseline methods with the same settings as our method using Mv-VGG on KinFaceW-I, KinFaceW-II, TSKinFace and Cornell datasets.

Features	KinFaceW-I					KinFaceW-II					TSKinFace					Cornell
	F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean	Mean
SILD	73.13	68.99	77.99	79.86	74.99	72.60	68.20	73.60	74.00	72.10	70.52	70.27	70.47	74.59	71.46	75.30
XQDA	73.35	72.35	77.59	78.35	75.41	80.20	76.60	78.20	80.00	78.75	70.82	70.82	74.46	77.38	73.37	76.96
MSIDA	77.86	74.81	79.67	81.84	78.54	75.80	69.60	74.60	76.60	74.15	74.75	74.35	74.66	75.09	74.71	81.16
MSIDA+WCCN	84.69	83.33	88.29	86.63	85.73	90.00	83.00	88.00	87.20	87.05	84.46	82.77	87.47	87.89	85.64	83.87
TXQDA	75.77	73.33	78.48	77.88	76.36	78.40	72.20	76.40	76.40	75.95	87.08	84.45	88.25	88.69	87.11	92.70
TXQEDA+WCCN	**91.00**	**87.78**	**92.32**	**93.35**	**91.11**	**89.80**	**90.60**	**87.60**	**93.20**	**90.30**	**89.42**	**89.31**	**90.87**	**93.15**	**90.68**	**93.77**

Table 10 shows a comparison of some baseline methods with the same settings. It shows a significant difference in performance by fusing several deep features (Mv-VGG). Our work beats the linear methods XQDA [27] by a large margin of 15%, 11%, 17% and 16%, while the margin for SILD [5] was 11%, 18%, 19% and 18% on KinFaceW-I, KinFaceW- II, TSKinFace and Cornell database, respectively. The improvement over TXQDA multi-linear subspace methods was 14.75%, 14.35%, 3.57% and 1.07%, while the improvement over MSIDA [8] was 12.57%, 16.15%, 15.97% and 12.61% on KinFaceW-I, KinFaceW-II, TSKinFace and Cornell databases, respectively. Otherwise, our work is better than the MSIDA+WCCN [8] approach with a performance of 5.38%, 3.25%, 7.51% and 9.90% for the KinFaceW-I, KinFaceW- II, TSKinFace and Cornell database, respectively. Since the proposed framework is based on several components that have higher performance, we consider the three components: the tensor mode, the WCCN technique, and the exponential criterion. The performance of the TXQEDA+WCCN approach using different component combinations compared to some baseline methods is shown in Table 10. Here, the best configuration is the proposed framework. We can clearly conclude that the combination of all components, that is: the tensor mode, the WCCN technique and the exponential criterion, led to the best performance. The results in Table 10 can confirm the statements mentioned in our work: (a) The multi-linear transformation (TXQEDA+WCCN) preserves the natural data structure by stacking the feature vectors in a tensor

representation, which provides more discriminative information. While linear transformation leads to loss of natural structure data when feature vectors are arranged in a linear manner. (b) Embedding WCCN metric in TXQEDA approach helps to solve the intra-class variability problem. (c) The samples are well distributed by the exponential criterion, where the pairs of the same family are folded into one point, maximally reducing the intra-class dispersion while increasing the relative separation between classes.

5.4. Computational Cost

The experiments were performed on a Windows 10 computer (Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10 GHz 2.10 GHz and 48GB RAM) using MATLAB 2018b. Table 11 shows the CPU time taken to test the kin relation between a pair of face samples using our TXQEDA+WCCN approach and the baseline TXQDA approach. The CPU time of feature extraction and total time are given in seconds (s), while the projection and matching time is given in milliseconds (ms). From Table 11, it can be seen that the time cost for projection and matching is negligible compared to the time for feature extraction. The total time cost for testing kinship faces with Mv-VGG features and TXQEDA+WCCN method is 1.08 (s). The test time is very small (not more than 0.0798 ms), so the proposed kinship verification method performs in real time. We can also find that the total time cost of TXQEDA+WCCN and TXQDA are almost identical, but our TXQEDA+WCCN method performs better than TXQDA as explained in this paper.

Table 11: Computational cost of our TXQEDA+WCCN for a pair of face images (parent-child) compared to the baseline TXQDA approach. The CPU feature extraction time and total time cost in seconds(s), while the projection and matching CPU time in milliseconds(ms)

Database	Feature extraction (s)	Projection (ms)		Matching (ms)		Total (s)	
	Mv-VGG	TXQEDA+WCCN	TXQDA	TXQEDA+WCCN	TXQDA	TXQEDA+WCCN	TXQDA
KinFaceW-II	1.0843	1.6000	0.9969	0.0798	0.0458	1.0860	1.0853
KinFac.eW-I		4.3548	1.6129	0.2758	0.0137	1.0889	1.0859
TSKinFace		2.6699	4.1748	0.1442	0.2350	1.0871	1.0887
Cornell KinFace		5.1852	6.8519	0.2765	0.4022	1.0898	1.0916

6. Conclusion

We proposed an efficient knowledge-based tensor system of for kinship verification based on multiple deep features. For feature extraction, we studied 4 well-known deep features: VGG-Face, VGG-M, VGG-S and VGG-F. To improve the robustness of the designed tensor and reduce its dimension, we proposed a Tensor Cross-view Quadratic Exponential Discriminant Analysis (TXQEDA) that integrates the Within-Class Covariance Normalization technique (WCCN). The WCCN metric can solve the within-class variability at each level and enhances the TXQEDA classification performance. The limitations of the proposed method (TXQEDA+WCCN) based on tensor analysis are mathematical complexity and coding difficulties. Despite these difficulties, the proposed framework illustrates the complementarity between the four deep feature descriptors in tensor subspace. The proposed knowledge-based tensor approach achieved successful results in kinship verification on four public databases KinFaceW-I, KinFaceW-II, TSKinFace and Cornell KinFace. In terms of perspectives, future work mainly involves two aspects. First, we envision extracting more discriminative features using less pre-trained networks. Our goal is to reduce the size of the training data while maintaining the peak performance already achieved. Second, we are interested in increasing the dissimilarity between negative pairs (the problem of classifying negative pairs as positive pairs) for the kinship verification task, since we have handled intra-class variability well in our work.

References

[1] J. Lu, X. Zhou, Y. 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.

[2] 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.

[3] X. Zhou, K. Jin, M. Xu, G. Guo, Learning deep compact similarity metric for kinship verification from face images, Information Fusion 48 (2019) 84 – 94. doi:https://doi.org/10.1016/j.inffus.2018.07.011.
URL https://www.sciencedirect.com/science/article/pii/S1566253517307273

[4] F. Dornaika, I. Arganda-Carreras, O. Serradilla, Transfer learning and feature fusion for kinship verification, Neural Computing and Applications (2019) 1–13.

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

[6] O. Laiadi, A. Ouamane, A. Benakcha, A. Taleb-Ahmed, A. Hadid, Kinship verification based deep and tensor features through extreme learning machine, in: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), 2019, pp. 1–4. doi:10.1109/FG.2019.8756627.

[7] J. Liang, Q. Hu, C. Dang, W. Zuo, Weighted graph embedding-based metric learning for kinship verification, IEEE Transactions on Image Processing 28 (3) (2019) 1149–1162. doi:10.1109/TIP.2018.2875346.

[8] O. Laiadi, A. Ouamane, A. Benakcha, A. Taleb-Ahmed, A. Hadid, Multi-view deep features for robust facial kinship verification, in: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 877–881. doi:10.1109/FG47880.2020.00118.

[9] X. Chen, X. Zhu, S. Zheng, T. Zheng, F. Zhang, Semi-coupled synthesis and analysis dictionary pair learning for kinship verification, IEEE Transactions on Circuits and Systems for Video Technology (2020) 1–1doi:10.1109/TCSVT.2020.3017683.

[10] M. Bessaoudi, A. Ouamane, M. Belahcene, A. Chouchane, E. Boutellaa, S. Bourennane, Multilinear side-information based discriminant analysis for face and kinship verification in the wild, Neurocomputing 329 (2019) 267 – 278. doi:https://doi.org/10.1016/j.neucom.2018.09.051.
URL https://www.sciencedirect.com/science/article/pii/S092523121831124X

[11] O. Laiadi, A. Ouamane, A. Benakcha, A. Taleb-Ahmed, A. Hadid, Tensor cross-view quadratic discriminant analysis for kinship verification in the wild, Neurocomputing 377 (2020) 286 – 300. doi:https://doi.org/10.1016/j.neucom.2019.10.055.
URL https://www.sciencedirect.com/science/article/pii/S0925231219314353

[12] R.-X. Hu, W. Jia, D.-S. Huang, Y.-K. Lei, Maximum margin criterion with tensor representation, Neurocomputing 73 (10) (2010) 1541 – 1549, subspace Learning / Selected papers from the European Symposium on Time Series Prediction. doi:https://doi.org/10.1016/j.neucom.2009.11.036.
URL https://www.sciencedirect.com/science/article/pii/S0925231210001128

[13] A. Chouchane, A. Ouamane, E. Boutellaa, M. Belahcene, S. Bourennane, 3d face verification across pose based on euler rotation and tensors, Multimedia Tools and Applications 77 (16) (2018) 20697–20714.

[14] X. Sun, M. Lv, Facial expression recognition based on a hybrid model combining deep and shallow features, Cognitive Computation 11 (4) (2019) 587–597.

[15] R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, International Journal of Computer Vision 126 (2) (2018) 144–157.

[16] M. Duan, K. Li, C. Yang, K. Li, A hybrid deep learning cnnelm for age and gender classification, Neurocomputing 275 (2018) 448 – 461. doi:https://doi.org/10.1016/j.neucom.2017.08.062.
URL https://www.sciencedirect.com/science/article/pii/S0925231217314923

[17] M. Tavakolian, A. Hadid, A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics, International Journal of Computer Vision 127 (10) (2019) 1413–1425.

[18] S. Singh, M. T. Ribeiro, C. Guestrin, Programs as black-box explanations, arXiv preprint arXiv:1611.07579 (2016).

[19] A. Blanco-Justicia, J. Domingo-Ferrer, Machine learning explainability through comprehensible decision trees, in: A. Holzinger, P. Kieseberg, A. M. Tjoa, E. Weippl (Eds.), Machine Learning and Knowledge Extraction, Springer International Publishing, Cham, 2019, pp. 15–26.

[20] A. Blanco-Justicia, J. Domingo-Ferrer, S. Martnez, D. Snchez, Machine learning explainability via microaggregation and shallow decision trees, Knowledge-Based Systems (2020) 105532.

[21] M. Wang, W. Deng, Deep face recognition: A survey, Neurocomputing 429 (2021) 215–244. doi:https://doi.org/10.1016/j.neucom.2020.10.081.
URL https://www.sciencedirect.com/science/article/pii/S0925231220316945

[22] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531 (2014).

[23] O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition (2015).

[24] 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.

[25] W. Yu, C. Zhao, Sparse exponential discriminant analysis and its application to fault diagnosis, IEEE Transactions on Industrial Electronics 65 (7) (2018) 5931–5940. doi:10.1109/TIE.2017.2782232.

[26] W. Wei, H. Dai, W. tai Liang, Exponential sparsity preserving projection with applications to image recognition, Pattern Recognition 104 (2020) 107357. doi:https://doi.org/10.1016/j.patcog.2020.107357.
URL https://www.sciencedirect.com/science/article/pii/S0031320320301606

[27] S. Liao, Y. Hu, X. Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[28] M. Vasilescu, D. Terzopoulos, Multilinear subspace analysis of image ensembles, in: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., Vol. 2, 2003, pp. II–93. doi:10.1109/CVPR.2003.1211457.

[29] 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

[30] L. Wolf, T. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: CVPR 2011, 2011, pp. 529–534. doi:10.1109/CVPR.2011.5995566.

[31] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, P. Ouellet, Front-end fac-

tor analysis for speaker verification, IEEE Transactions on Audio, Speech, and Language Processing 19 (4) (2011) 788–798.

[32] H. Yu, C. Y. Chung, K. P. Wong, H. W. Lee, J. H. Zhang, Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition, IEEE Transactions on Power Systems 24 (2) (2009) 661–667.

[33] X. Qin, X. Tan, S. Chen, Tri-subject kinship verification: Understanding the core of a family, IEEE Transactions on Multimedia 17 (10) (2015) 1855–1867.

[34] 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.

[35] 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.

[36] J. Lu, J. Hu, Y. 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.

[37] O. Laiadi, A. Ouamane, E. Boutellaa, A. Benakcha, A. Taleb-Ahmed, A. Hadid, Kinship verification from face images in discriminative subspaces of color components, Multimedia Tools and Applications 78 (12) (2019) 16465–16487.

[38] A. Moujahid, F. Dornaika, A pyramid multi-level face descriptor: application to kinship verification, Multimedia Tools and Applications 78 (7) (2019) 9335–9354.

[39] X. Zhou, K. Jin, M. Xu, G. Guo, Learning deep compact similarity metric for kinship verification from face images, Information Fusion 48 (2019) 84–94.

[40] 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.

[41] N. Kohli, D. Yadav, M. Vatsa, R. Singh, A. Noore, Supervised mixed norm autoencoder for kinship verification in unconstrained videos, IEEE Transactions on Image Processing 28 (3) (2018) 1329–1341.

[42] L. Zhang, Q. Duan, D. Zhang, W. Jia, X. Wang, Advkin: Adversarial convolutional network for kinship verification, IEEE Transactions on Cybernetics (2020) 1–14doi:10.1109/TCYB.2019.2959403.

[43] H. Dibeklioglu, Visual transformation aided contrastive learning for video-based kinship verification, in: The IEEE International Conference on Computer Vision (ICCV), 2017.

[44] S. Wang, H. Yan, Discriminative sampling via deep reinforcement learning for kinship verification, Pattern Recognition Letters 138 (2020) 38 – 43. doi:https://doi.org/10.1016/j.patcog.2016.08.020.

[45] H. Yan, C. Song, Multi-scale deep relational reasoning for facial kinship verification, Pattern Recognition (2020) 107541.

[46] Y.-G. Zhao, Z. Song, F. Zheng, L. Shao, Learning a multiple kernel similarity metric for kinship verification, Information Sciences 430 (2018) 247–260.

[47] M. B. López, E. Boutellaa, A. Hadid, Comments on the kinship face in the wild data sets, IEEE transactions on pattern analysis and machine intelligence 38 (11) (2016) 2342–2344.