Kinship verification between two people by machine learning

Oualid Laiadi

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES HAUTS-DE-FRANCE

Université Polytechnique HAUTS-DE-FRANCE

UNIVERSITY OF BISKRA

order number: 21/08

PhD Thesis

Submitted for the degree of Doctor of Philosophy from Polytechnic University of Hauts-de-France (UPHF), INSA HAUTS-DE-FRANCE and, University of Mohamed Khider Biskra (UMKB)

Subject

Kinship verification between two people by machine learning

Presented and defended by **Oualid Laiadi**

On: 13/09/2021, Biskra

Doctoral school:

Doctoral School Polytechnique Hauts-de-France (ED PHF)

Thesis directors:

Mr. Abdelmalik Taleb-Ahmed	Professor	University of Hauts-de-France (France)	Co-director
Mr. Abdelhamid Benakcha	Professor	University of Biskra (Algeria)	Co-director

Jury:

Ms. Malika Mimi	Professor	University of Mostaganem (Algeria)	President
Ms. Nacéra Benamrane	Professor	University of Oran (Algeria)	Reviewer
Mr. Kamel Eddine Melkemi	Professor	University of Batna (Algeria)	Reviewer
Mr. Amir Nakib	MC HDR	University of Créteil (France)	Reviewer
Ms. Salima Ouadfel	MC HDR	University of Constantine (Algeria)	Examiner
Mr. Abdeljalil Ouahabi	Professor	University of Tours (France)	Examiner
Ms. Atika Rivenq	Professor	University of Hauts-de-France (France)	Invited
Mr. Abdenour Hadid	Professor	University of Oulu (Finland)	Invited

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES HAUTS-DE-FRANCE

Université Polytechnique HAUTS-DE-FRANCE

UNIVERSITY OF BISKRA

numéro d'ordre: 21/08

Thèse de doctorat

Pour obtenir le grade de Docteur de l'UNIVERSITÉ POLYTECHNIQUE HAUTS-DE-FRANCE, l'INSA HAUTS-DE-FRANCE et l'UNIVERSITÉ MOHAMMED KHIDER DE BISKRA

Sujet

Vérification de la parenté entre deux personnes par apprentissage machine

Présentée et soutenue par **Oualid Laiadi**

Le: 13/09/2021, Biskra

Ecole doctorale:

École Doctorale Polytechnique Hauts-de-France (ED PHF)

Directeur de thèse:

Mr. Abdelmalik Taleb-Ahmed	Professeur	Université de Hauts-de-France (France)	Co-directeur
Mr. Abdelhamid Benakcha	Professeur	Université de Biskra (Algeria)	Co-directeur

Jury:

Ms. Malika Mimi	Professeur	Université de Mostaganem (Algérie)	Présidente
Ms. Nacéra Benamrane	Professeur	Université d'Oran (Algérie)	Rapporteurs
Mr. Kamel Eddine Melkemi	Professeur	Université de Batna (Algérie)	Rapporteurs
Mr. Amir Nakib	MC HDR	Université de Créteil (France)	Rapporteurs
Ms. Salima Ouadfel	MC HDR	Université de Constantine (Algérie)	Examinateurs
Mr. Abdeljalil Ouahabi	Professeur	Université de Tours (France)	Examinateurs
Ms. Atika Rivenq	Professeur	Université de Hauts-de-France (France)	Invité
Mr. Abdenour Hadid	Professeur	Université d'Oulu (Finland)	Invité

Research team, Laboratory:

Institute of Electronics Microelectronics and Nanotechnology - Department of Opto-Acousto-Electronics (IEMN DOAE - UMR 8520)

and

Laboratory of Expert Systems, Imaging and their Applications in engineering (LESIA), University of Mohamed Khider Biskra (UMKB), Algeria.

"Science (or magnanimity) are bitter-tasting at first, but sweeter than honey in the end".
Dish with epigraphic decoration, Louvre Museum.

Every living being is an engine geared to the wheelwork of the universe. Though seemingly affected only by its immediate surrounding, the sphere of external influence extends to infinite distance.

Nikola Tesla

To see a world in a grain of sand and a heaven in a wild flower, hold infinity in the palm of your hand and eternity in an hour.

William Blake

If you cannot do great things, do small things in a great way.

Napoleon Hill

If opportunity doesn't knock, build a door.

Milton Berle

Imagination is everything. It is the preview of life's coming attractions.

Albert Einstein

Opportunity is missed by most people because it is dressed in overalls and looks like work.

Thomas Edison

Two roads diverged in a wood, and I—I took the one less traveled by, and that has made all the difference.

Robert Frost

I dedicate this work to my parents who have always believed in me and supported me

To my Mother and Father

To my Sisters and Brothers

Oualid Laiadi,

Acknowledgments

The work of this thesis is the result of my research and exchanges with many people during the three long and short years at UPHF, UMKB, IEMN, and LESIA. These years have been very rewarding both professionally and personally. The first person I want to thank is the thesis supervisors, Mr. Abdelmalik TALEB-AHMED and Mr. Abdelhamid BENAKCHA. They guided me during this thesis and gave me advice and encouragement to keep going. I am really grateful for their moral support, for not stressing me at all, for their patience and understanding especially when I had so many questions, and for making me believe more in myself and my work. Working with them enriched me with lessons about success, problem-solving, positive thinking, critical thinking, risk-taking, progress, and enthusiasm. It is a pleasure to work with such people. Personally, I think they are the best supervisors you can get.

I am also thankful for Mr. Abdenour HADID and Mr. Abdelmalik OUAMANE for being the advisors of my thesis for three years and for their readiness to help me through the work of this thesis. Their encouragement and advice have helped me a lot especially their advice about the writing of this report. I am really grateful for their kindness and support and it was a pleasure to be directed by them.

A special thanks go to Ms. Atika RIVENQ, for her support and help during pursuing my Ph.D. at University Polytechnique Hauts-de-France.

A particular thanks go to professors, Ms. Malika MIMI, Ms. Nacéra BENAMRANE, Mr. Kamel Eddine MELKEMI, Mr. Amir NAKIB, Ms. Salima OUADFEL, and Mr. Abdeljalil OUAHABI, the jury members for having agreed to review and evaluate this work.

Finally, I thank my family and my friends for their patience, encouragement, and support which were very useful to me during my work.

Abstract

The kinship verification field attracted much attention in the few past years due to its capacity to improve biometrics systems as a soft biometric for face verification (kinship traits) and an important role in many society applications (kinship verification). Among these applications include the creation of family trees, family album organization, image annotation, finding missing children and forensics. Although, the DNA test is the most trustworthy way for kinship verification, it cannot be used in many situations. Automatic kinship verification from facial images can exemplary be done in video surveillance scenes.

In this thesis, facial kinship verification over facial images is studied. At this end, we start with the previously proposed approaches like features learning-based kinship verification methods, metric learning-based kinship verification methods, and convolutional deep learning-based kinship verification methods. Also, the general facial kinship verification system is presented, challenges and measures of characteristics are mentioned. Furthermore, the various evaluation terms are illustrated. Concluding with the proposed approaches and the obtained results on various databases. The proposed frameworks comprise of three main phases as follows: 1) features extractions; 2) subspace transformations analysis; 3) kinship verification decision.

The aim of feature extraction is to extract discriminative representations of facial images. This phase is important since the kinship traits are very sensitive to the unconstrained environments (i.e. facial images captured under uncontrolled environments without any restrictions in terms of pose, lighting, background, expression, and partial occlusion). Also, it can affect the final decision performance of the framework. Subspace transformations analysis phase extract and select the more attractive and discriminative facial traits. Therefore, the features are extracted by a projection of the original data (features) of the previous phase to get better discrimination and make more precise decisions. In the last phase, cosine similarity is used as the best metric compatible with discriminant analysis methods (subspace transformations analysis methods) and kinship verification. The final metric between two facial images is compared to a threshold to decide if the pair facial images come from the same family or not.

Finally, our results show great improvement for facial kinship verification on the largest and smallest databases. Also, a robust and good performance was achieved by the proposed systems and comparing favorably with the state of the art approaches. The proposed frameworks are also convenient for real-time applications.

Keywords: facial kinship verification, facial images, feature extraction, subspace transformations analysis, features learning-based kinship verification, metric learning-based kinship verification, convolutional deep learning-based kinship verification, forensics.

Résumé

Le domaine de la vérification de la parenté a attiré beaucoup d'attention ces dernières années en raison de sa capacité à améliorer les systèmes biométriques en tant que biométrique souple pour la vérification du visage (traits de parenté) et un rôle important dans de nombreuses applications de la société (vérification de la parenté). Parmi ces applications, citons la création d'arbres généalogiques, l'organisation d'albums de famille, l'annotation d'images, la recherche d'enfants disparus et la criminalistique. Bien qu'un Le test ADN est le moyen le plus fiable pour la vérification de la parenté, il ne peut pas être utilisé dans de nombreuses situations. La vérification automatique de la parenté à partir d'images faciales peut être réalisée à titre d'exemple dans les scènes de vidéosurveillance.

Dans cette thèse, la vérification de la parenté faciale sur les images faciales est étudiée. À cette fin, nous commençons avec les approches précédemment proposées telles que les méthodes de vérification de parenté basées sur les fonctionnalités, les méthodes de vérification de parenté basées sur l'apprentissage métrique et les méthodes de vérification de parenté basées sur l'apprentissage profond par convolution. En outre, le système général de vérification de la parenté faciale est présenté, les défis et les mesures des caractéristiques sont mentionnés. De plus, les différents termes d'évaluation sont illustrés. Conclusion avec les approches proposées et les résultats obtenus sur diverses bases de données. Les systèmes proposés comprennent trois phases principales comme suit: 1) extractions de caractéristiques; 2) analyse des transformations du sous-espace; 3) décision de vérification de la parenté.

Le but de l'extraction de traits est d'extraire des représentations discriminantes d'images faciales. Cette phase est importante car les traits de parenté sont très sensibles aux environnements non contraints (i.e. images capturées dans des environnements non contrôlés sans aucune restriction termes de pose, d'éclairage, d'arrière-plan, d'expression et d'occlusion partielle). En outre, cela peut affecter la performance de décision finale du système. La phase d'analyse des transformations du sous-espace extrait et sélectionne les traits du visage les plus attrayants et discriminants. Par conséquent, la les caractéristiques sont extraites par une projection des données originales (caractéristiques) de la phase précédente pour obtenir une meilleure discrimination et prendre des décisions plus précises. Dans la dernière phase, la similarité cosinus est utilisée comme la meilleure métrique compatible avec les méthodes d'analyse discriminante (méthodes d'analyse des transformations de sous-espace) et la vérification de parenté. La métrique finale entre deux images faciales est comparée à un seuil pour décider si les images faciales de la paire proviennent ou non de la même famille.

Enfin, nos résultats montrent une grande amélioration pour la vérification de la parenté faciale sur les bases de données les plus grandes et les plus petites. En outre, les systèmes proposés ont obtenu une performance robuste et bonne et se comparent favorablement l'état

de l'art approche. Les systèmes proposés sont également pratiques pour les applications en temps réel.

Mots clés: vérification de la parenté faciale, images faciales, extraction de caractéristiques, analyse des transformations du sous-espace, vérification de la parenté basée sur les caractéristiques, vérification de la parenté basée sur l'apprentissage métrique, Vérification de la parenté basée sur l'apprentissage profond par convolution, criminalistique.

Contents

List of Acronyms

- **3DMDA** Three-dimensional Modular Discriminant Analysis
- **BRIEF** Binary Robust Independent Elementary Features
- **BSIF** Binarized Statistical Image Features
- **CIF** Common Intermediate Format
- **CNN** Convolutional Neural Network
- **CoALBP** Co-occurrence of Adjacent Local Binary Patterns
- **D&S** Deep&Shallow
- **DDMML** Discriminative Deep Multi-Metric Learning
- **DMML** Discriminative Multi-Metric Learning
- **DNA** Deoxyribonucleic Acid
- **DNN** Deep Neural Network
- **EDA** Exponential Discriminant Analysis
- **EER** Equals Error Rate
- **F-D** Father-Daughter
- **F-S** Father-Son
- **FAR** False Accept Rate
- **FIW** Families In the Wild
- **FLD** Fisher's Linear Discriminant
- **FM-D** Father-Mother-Daughter
- **FM-S** Father-Mother-Son
- **FRR** False Reject Rate
- **GTDA** General Tensor Discriminant Analysis

- **HOG** Histogram of Oriented Gradients
- **HSL** Heterogeneous Similarity Learning
- **ICA** Independent Component Analysis
- **KISSME** Keep It Simple and Straightforward MEtric
- **KML** Kinship Metric Learning
- **KNN** K-Nearest Neighbors
- **KSVM** Kernel Support Vector Machine
- **LBP** Local Binary Pattern
- **LDA** Linear Discriminant Analysis
- **LFW** Labeled Faces in the Wild
- **LPQ** Local Phase Quantization
- **LR** Logistic Regression
- **M-D** Mother-Daughter
- **M-S** Mother-Son
- **MDA** Multilinear Discriminant Analysis
- **MKSM** Multiple Kernel Similarity Metric
- **MNRML** Multiview Neighborhood Repulsed Metric Learning
- **MPCA** Multilinear principal Component Analysis
- **MSDA** Multilinear Spatial Discriminant Analysis
- **MSIDA** Multilinear Side-Information based Discriminant Analysis
- **MvDML** Multi-view Discriminative metric learning
- **NRCML** Neighborhood Repulsed Correlation Metric Learning
- **NRML** Neighborhood Repulsed Metric Learning
- **PCA** Principal Component Analysis
- **PDFL** Prototype Discriminative Feature Learning

- **RBF** Radial Basis Function
- **ROC** Receiver Operating Characteristic
- **SIEDA** Side-Information based Exponential Discriminant Analysis
- **SIFT** Scale-Invariant Feature Transform
- **SILD** Side-Information based Linear Discriminant analysis
- **SIWEDA** Side Information Weighted Exponential Discriminant Analysis
- **SOA** State Of the Art
- **SPLE** Spatial Pyramid LEarning
- **STD** Standard Deviation error
- **STDA** Sparse Tensor Discriminant Analysis
- **SURF** Speeded Up Robust Features
- **SVM** Support Vector Machine
- **StatBIF** Statistical Binarized Image Feature
- **TLBP** Three-patch LBP
- **TXQDA** Tensor Cross-view Quadratic Discriminant Analysis
- **UB** University of Buffalo
- **UMDA** Uncorrelated Multilinear Discriminant Analysis
- **VGG** Visual Geometry Group
- **VR** Verification Rate
- **WCCN** Within Class Covariance Normalization
- **XQDA** Cross-view Quadratic Discriminant Analysis
- **YTF** YouTube Face

List of Figures

List of Tables

1 Introduction

Contents

1.1 Context and motivation

The capacity to determine the identity of persons and to accord personal traits (e.g. name, age, nationality, and so on) with a person has been very intrinsic to the structure of our society. Generally, humans utilized appearance characteristics such as voice, face and gait as well as other contextual data information (for example, clothing and location) to identify themselves. The set of traits associated with an individual describes their own personal identity. At the beginning of civilization, people lived in limited small communities where persons could easily identify each other. Furthermore, a big explosion of population expansion accompanied by raised mobility in the modern society which has required the development of advanced identity management automatic-based systems that can efficiently record, preserve and erase the private identities of peoples.

A facial biometric system is subdivided into two phases, the training stage (offline training) and the test stage (online classification/verification). The training stage will be carried out only once in which the enrollment of the facial images of the different individuals is used in order to extract and describe the biometric signature of each individual. During the test stage, the new data is compared with the training data that is automatic learned in the training stage in order to make a decision to accept or reject the candidate. The steps carried out in these two phases in a face recognition system are subdivided into three main modules [70]: face detection, feature extraction and recognition (classification). The general face recognition system scheme is illustrated in Fig. 1.1. A detailed face recognition system scheme is illustrated in Fig. 1.2.

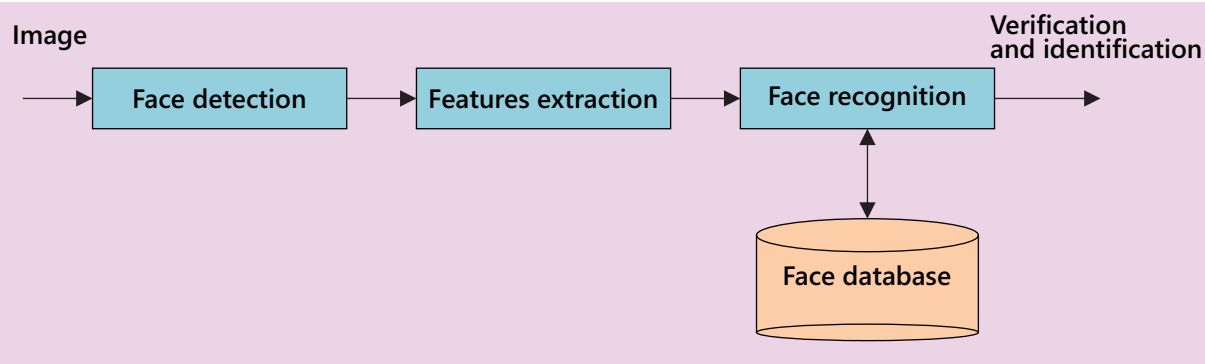


Figure 1.1: Face recognition structure.

Three essential steps are utilized to develop a robust facial recognition system: (i) face detection, (ii) feature extraction, and (iii) face recognition (shown in Figure 1.1). The face detection stage is utilized to detect and determine the human facial image acquired by the system. The feature extraction stage is used to extract the discriminant feature data for each human face determined in the first stage. Finally, the face recognition stage comprises the extracted features from the human facial that compare it with whole template facial databases to determine the human facial identity.

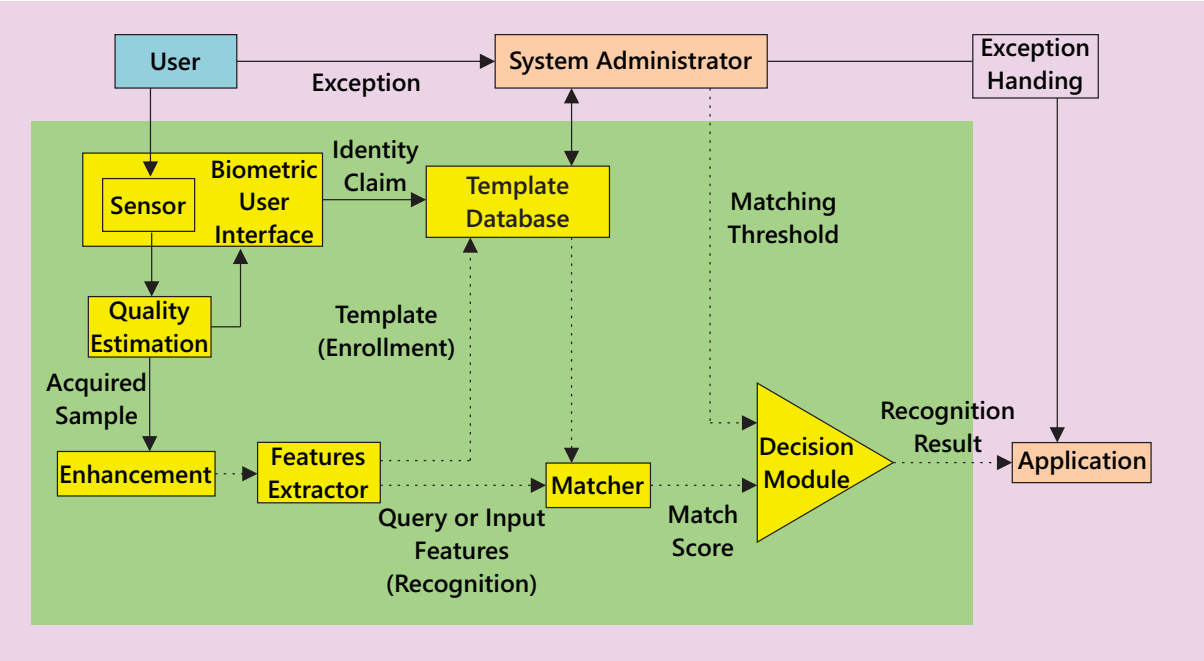


Figure 1.2: Basic building blocks of a generic biometric system [72].

- **Face Detection:** The face recognition system starts first with the center on of the human facials in a input image. The aim of this stage is to define if the target image comprises human facials or not. The illumination variations and facial expression can block proper facial detection. In order to make easier the design of a robust face recognition system and create it more effective, pre-processing stages are performed. Many approaches are utilized to detect and define the human facial image, as an example, Viola–Jones detector [152,180], histogram of oriented gradient (HOG) [137], and principal component analysis (PCA) [141]. Furthermore, the face detection stage can be utilized for image and video classification [122], regression [51], object tracking [148], region-of-interest detection [141], and so on.

- **Feature Extraction:** The essential function of this stage is to describe the facial images captured in the detection stage. This stage explains a face as a group of features vector called a "traits" that characterizes the discriminate features of the facial image such as mouth, nose, and eyes with their geometrical distribution [122]. Each facial is described by its size, structure, and shape, which permit it to be identified. Many approaches involve extracting the form/shape of the eyes, mouth, or nose to identify the facial utilizing the size and/or distance [122]. HOG [137], Eigenface [149], independent component analysis (ICA) [88], Gabor filter [108] approaches are widely utilized to extract the facial features.

- **Face Recognition:** This stage considers the features vector extracted from the background within the feature extraction stage and compares it with similar facial stocked in a specific dataset. There are two essential general applications for face

recognition, one is called recognition or identification and another one is called verification. Within the identification stage, a test facial is compared with a group of facials aiming to find the most similar match. Within the verification stage, a test facial is compared with a known facial in the dataset in order to make a decision of the acceptance or rejection.

Several systems proposed and implemented to identify a human facial in 2D or 3D images. We classify these systems into three methods based on their detection and recognition approach (Fig. 1.3): (1) local, (2) holistic (subspace), and (3) hybrid approaches. The first method is classified according to specific face features, not considering the whole face. The second method employs the whole face as input information and then projects into a small and discriminative subspace or in correlation sub-plane. The third method utilizes global and local features in order to enhance face recognition accuracy.

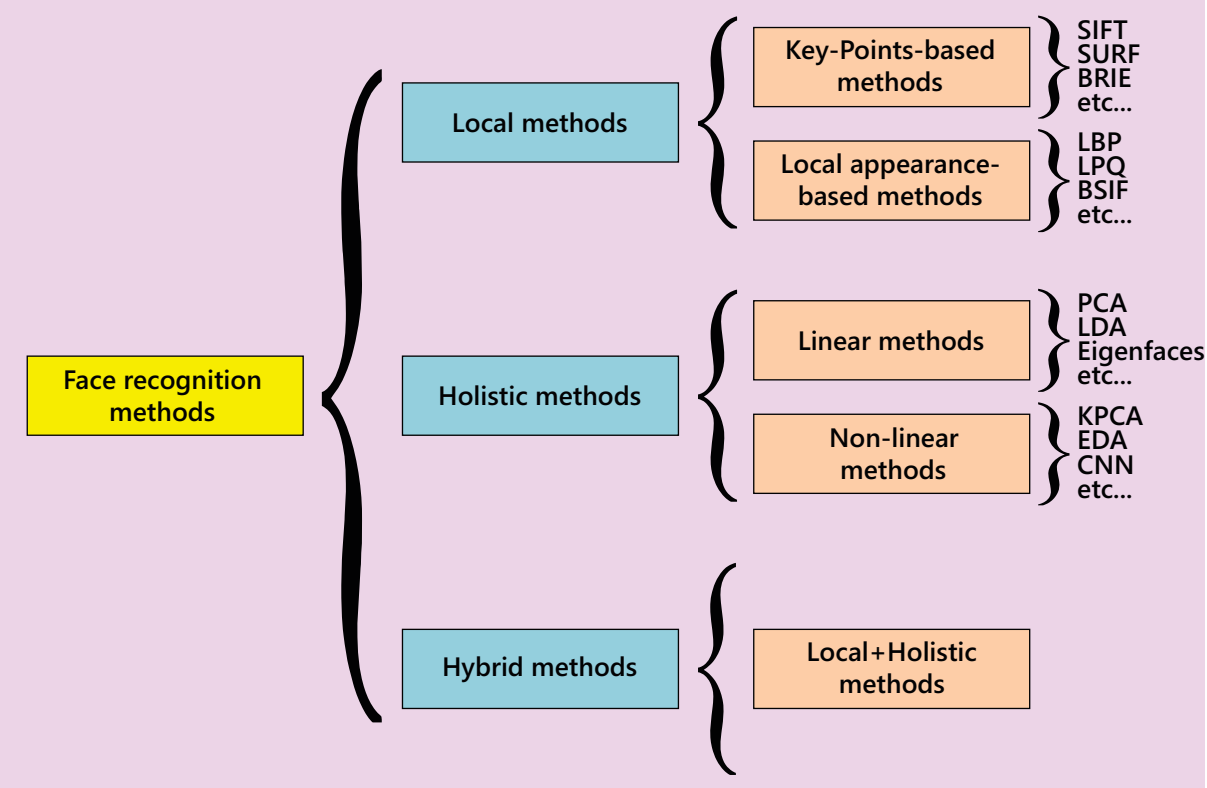


Figure 1.3: Face recognition methods. SIFT, scale-invariant feature transform; SURF, Speeded Up Robust Features; BRIEF, binary robust independent elementary features; LBP, local binary pattern; BSIF, binarized statistical image feature; LPQ, local phase quantization; PCA, principal component analysis; LDA, linear discriminant analysis; KPCA, kernel PCA; CNN, convolutional neural network; EDA, exponential discriminant analysis.

Kinship verification from face images, one of the new topics in computer vision that has been studied and used for several years, can be applied to potential applications, such as the creation of family trees, family album organization, image annotation, finding missing children and forensics. Checking if two persons are from the same family or not can be automatically verified through facial images. Learn and extract the face similarities between family members is challenging. Many encouraging results have been shown over

the past a few years, kinship verification from face images still remains open. Although, a DNA test is the most trustworthy way for kinship verification, it cannot be used in many situations. Automatic kinship verification from facial images can exemplary be done in video surveillance scenes. In addition to the obstacle generally faced the face verification in unconstrained environments (i.e. facial images captured under uncontrolled environments without any restrictions in terms of pose, lighting, background, expression, and partial occlusion), kinship verification inserts another layer of obstacles which is far from being easy. Kinship verification treats facial images which belong inevitably to different persons with a considerable age difference and in some condition with different gender. Further, the face traits of persons of the same family may offer a large dissimilarity whereas pair faces of persons with no kinship may look similar. All these challenges greatly increase the difficulties of the automatic kinship verification problem.

Through the different chapters, we highlight the interest of using algorithms based on mono-dimensional (vector-based) and multi-dimensional (tensor-based) analysis using deep and shallow features in kinship verification.

1.2 Kinship verification challenges

Kinship verification, one of the basic topics in computer vision and pattern recognition, has received substantial attention in recent years. Many approaches have been proposed to kinship verification in unconstrained environments, while each of these approaches consists in checking if two persons are from the same family or not through facial images.

Verifying kinship through facial images is difficult due to the high degree of variability of the visible effects such as genetic difference, gender difference and age gap. In short, the following two factors have a major impact on problem solving:

- **Unique challenges:** The appearance gap in the kinship verification problem is much greater than in the traditional facial verification setup (for example, by looking at two pictures with different genders and different ages, and checking whether these two subjects have a relationship between the parent and child). Moreover, the relationships of different relatives will have different patterns of similarity. These may cause major challenges for all facial kinship verification frameworks.

- **Common challenges:** Due to the challenges in verifying faces, the appearance of close-up faces is sensitive to changing various factors, such as variations in facial expressions, obstruction and position. Furthermore, some other influence factors may be presented in the real scene, like to as illumination, opacity or low resolution, may change the visual representation of kinship from facials in various ways.

Figure 1.4 illustrates the mentioned challenges. Several algorithms have been suggested to meet these challenges over the past decade. More recently, [4, 30], have studied various

representative approaches of verifying kinship in the situation of the small sample training data only with a few kinds of relations, namely father-son, father-daughter, mother-son and mother-daughter. Therefore, it remains to be test whether modern kinship checking approaches work well on large samples training data with various closer relationships, particularly now that the released FIW dataset has been published [138]. As Figure 1.4 illustrates, the more varied samples relationships of family members pose larger challenges to the issue of kinship verification and are far from resolved.

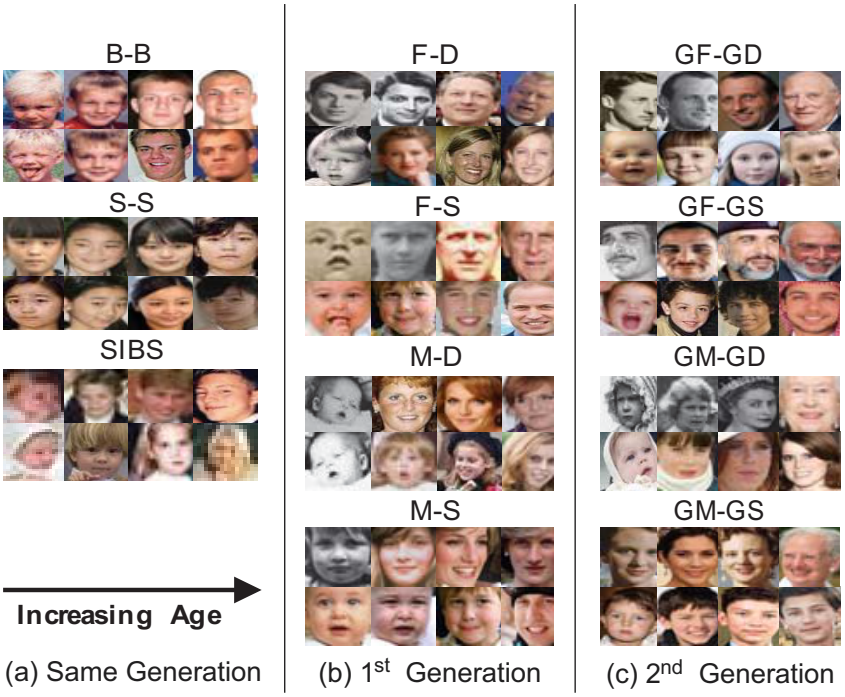


Figure 1.4: Samples of 11 pair types of FIW. Each type is of a unique pair randomly selected from a set of diverse families to show variation in ethnicity, while four faces of each individual depict age variations [138].

1.3 Benchmark databases

To evaluate the performance of the proposed kinship verification approaches, we considered six kinship databases: Cornell KinFace database, UB KinFace database, TSKinFace database, KinFaceW-I database, KinFaceW-II database and FIW database. These databases consist of four kinds of Parent-Child relationships (except FIW database which contains eleven relations of four Parent-Child relations, three Siblings relations and four Grandparent-Grandchild relations). The face images are with various ages and ethnicities, and captured under uncontrolled environments and no restriction in terms of pose. For face verification (face matching or self-kinship), we considered two challenging databases namely Labeled Faces in the Wild (LFW) database and YouTube Face (YTF) database.

1.3.1 Kinship verification databases

Cornell KinFace database [44] consists of 143 pairs of parents and children images gathered from the web. There are 286 cropped frontal face images of size 100×100 pixels. Most of the images were taken from Google Images. To ensure that the facial extracted characteristics are in high quality, only frontal face images with a neutral facial expression are chosen. We note that, 7 families are taken out of the original database which consists of 150 families for privacy issues.

UB KinFace database [170] includes 600 images of 400 people which are divided into 200 pairs of child-young parent (set 1) and 200 pairs of child-old parent (set 2). These two sets of pairs are used to enhance, test, and evaluate kinship verification algorithms. Most of images in the database are real-world combinations of public figures (celebrities and politicians) from Internet. It is the first database that comprises all children, young parents and old parents for the purpose of kinship verification.

TSKinFace database [133] consits of two types of tri-subject kinship relations which are: Father-Mother-Daughter (FM-D) and Father-Mother-Son (FM-S). The FM-D contains 502 relations and FM-S has 513 relations (4060 face images). These images are from public figures gathered from the Internet. The face images are cropped using the position of eyes into 64×64 pixels resolution. For fair comparison, we restructured the database by separating the group of Father-Mother-Daughter into two groups Father-Daughter and Mother-Daughter kinship relations, and the group of FatherMother-Son into two groups Father-Son and Mother-Son kinship relations.

Kinship Face in the Wild database (KinFaceW) [103] consists of two different sub-databases: KinFaceW-I and KinFaceW-II. Both sub-databases are gathered through Internet research, including some public figures with their parents and/or children. In the KinFaceW-I dataset, there are 156, 134, 116, and 127 pairs corresponding to the F-S, F-D, M-S, and M-D relations, respectively. For the KinFaceW-II dataset, each kin relation type contains 250 pairs. In total KinFaceW-I counts 1066 face images and 2000 face images for KinFaceW-II.

FIW database [138] we considered the largest FIW kinship database using: four relations, Grandfather-Granddaughter (GF-GD), Grandfather-Grandson (GF-GS), Grandmother-Granddaughter (GM-GD) and Grandmother-Grandson (GM-GS) face subsets. In GF-GD subset, there are 7,078 pairs of images for positive and negative relations. In GF-GS subset, there are 4,830 pairs of images for positive and negative relations. In GM-GD subset, there are 6,512 pairs of images for positive and negative relations. In GM-GS subset, there are 4,614 pairs of images for positive and negative relations.

1.3.2 Face verification databases

Labeled Faces in the Wild (LFW) database [67] is a big dataset collected from the web, specially gathered to study the problem of face recognition in unconstrained environments containing real-world variations in terms of lighting, pose, expressions, blur, occlusion, resolution, and so on. This challenging dataset consists 13,233 facial images belonging to 5,749 different subjects.

YouTube Face (YTF) database [162] consists of 3425 videos from 1595 different subjects with various variations of pose, expression and illumination, and the average length of each video clip is 181.3 frames.

1.4 Objectives and contributions

The main thesis focuses on the development, implementation and evaluation of automatic and efficient kinship verification frameworks based on metric learning techniques of linear and multi-linear subspaces in uncontrolled environments in which the variations of pose, lighting, background, expression, and partial occlusion are very different between the training and the test classes.

We can organize our contributions in classification stage into two essential categories: i) vector-based methods and ii) tensor-based methods. These two categories need the features extraction stage, which we categorized theme into two essential categories as it: i) shallow features (shape/texture features), and deep&shallow features. Figure 1.5 illustrated our main contributions on classification stage using different features extraction categories for kinship verification.

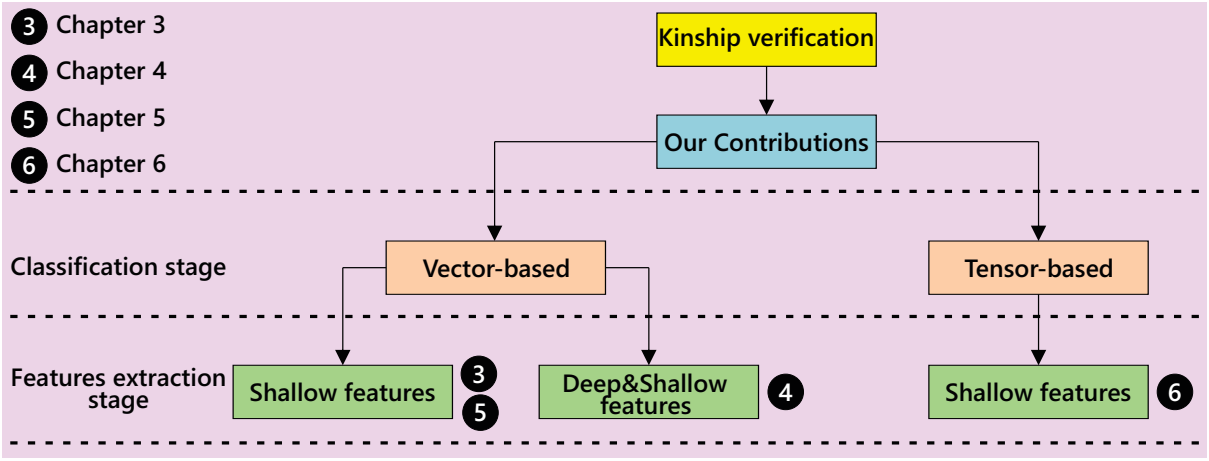


Figure 1.5: Thesis map: our main contributions over different categories of classification stage (vector-based and tensor-based strategies design) for kinship verification using different features extraction categories.

We summarize the main research contributions of this thesis as follows:

- Study of the State Of the Art (SOA) of different kinship verification approaches based on deep learning and metric learning.

- Develop and design of robust kinship verification frameworks against variations in expression, illumination and pose, based on vector-based and tensor-based metric learning, using the shallow features (i.e. texture/shape) and deep features of intensity facial images and color facial images.

We can subdivide our contributions into four folds as follow:

I- First contribution (illustrated in Chapter 3):

- We introduce an efficient method for facial kinship verification based on multiple scale feature extraction projected through Side Information Exponential Discriminant Analysis (SIEDA) subspace and combined different features using Logistic Regression (LR) method scores fusion.
- We evaluate the effectiveness of color-texture information data over discriminative subspace utilizing two-step learning technique, SIEDA and Logistic Regression, for automatic facial verification of kinship from facial images.
- We evaluate different color spaces and descriptors on four benchmark kinship databases. Especially, each color channel of face image from a specified color space is projected through the same implicit learned color channel subspace, and then all the channels information are combined to achieve better discrimination.
- We study the combination of the different descriptors from the different color components.

II- Second contribution (illustrated in Chapter 4):

- We introduce a novel discriminative subspace of the proposed Side-Information based Linear Discriminant analysis integrating Within Class Covariance Normalization (SILD+WCCN) subspace transformation analysis method for facial kinship verification. Therefore, WCCN minimises the class intra-variability impact by minimising the expected classification error on the training stage [9].
- We suggest a two robust automated facial kinship verification systems appropriate for bi-subject and tri-subject kinship verification, from facial images captured in unconstrained environments. The facial data is exemplified as a multiple view feature based on the fusion of different deep and shallow features in order to get a more discriminative facial model.
- We evaluate the effectiveness of deep/shallow information data over a novel discriminative subspace using the two-step learning technique, SILD+WCCN, and Logistic Regression, for automatic verification of kinship from facial images.

- We extensively test our SILD+WCCN/LR technique versus the state-of-the-art approaches utilizing two challenging facial kinship databases namely KinFaceW-II and TSKinFace.

III- Third contribution (illustrated in Chapter 5):

- Introducing a new native feature for describing facial photos. Our descriptor is based on the local statistical features of the face image and the original BSIF descriptor.
- We propose a new SIWEDA method to verify face and kinship based on the classic SIEDA method. Moreover, to mitigate the internal variance of the class, we proposed two variants SIEDA + WCCN and SIWEDA + WCCN by incorporating WCCN into SIEDA and SIWEDA, respectively.
- We broadly evaluate our approach against the state-of-the-art approaches using five challenging face and kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, YTF and LFW.

IV- Fourth contribution (illustrated in Chapter 6):

- For the first time, we are dealing with the problem of verifying facial kinship as a cross-view matching problem because each kinship usually changes from two facial images belonging to two different people.
- We suggest a robust and suitable automatic face verification framework for kinship verification, from face photos taken in unrestricted environments. The face data is represented as a high-level tensor that relies on a combination of different local features in order to provide a more robust face model.
- We propose a new method for reducing and classification dimensions, called Tensor Cross-view Quadratic Discriminant Analysis (TXQDA), which preserves data structure, expands the margin between samples, helps alleviate the problem of small sample size, and reduces computational cost.
- We evaluate our TXQDA method broadly against state-of-the-art methods using five challenging facial kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II, and FIW.

Finally, we can classified our contributions (cited in Chapters 3,4,5 and 6) as mentioned in Figure 1.5 as follows: In Chapter 3, we used the combination of Shallow features projected through SIEDA (vector-based) method. In Chapter 4, we used the combination of Deep&Shallow features projected through the proposed SILD+WCCN (vector-based) method. For Chapter 5, we used the combination of Shallow features projected through the proposed SIWEDA+WCCN (vector-based) method. For Chapter 6, we used the combination of Shallow features over a tensor design projected through the proposed TXQDA (tensor-based) method.

1.5 Machine learning explanations

In computer vision field, a published framework in the literature works should take into account several points and must be explainable. As mentioned in [112], an explanation way for a black-box (a framework/a method) machine learning approach should take into account the following properties:

• **Accuracy**. This trait refers to how degree of success that an explanation predicts new tested data (unseen data). weak explanation accuracy could be fine only if the black-box framework to be explained is also inaccurate.

• **Fidelity**. The explained model predictions should correspond and conclude the explanations. There is high relation between accuracy and fidelity: where if the explanation has larger fidelity and black-box model is high accurate, the explanation of the model has also larger accuracy

• **Consistency**. Explanations must apply equally important to all model trained using the same train data set.

• **Stability**. Similar instances must present similar explanations, as long as particular instances was provided its explanations.

• **Representativeness**. A highly representative explanation is one that can be applied to many decisions on many instances.

• **Certainty**. If the method at study provides a measure of confidence on its decisions, an explanation of this decision must reflect this.

• **Novelty**. This property indicates to the ability of the explanation paradigm to cover instances far from the training space.

• **Degree of importance**. The explanation must pinpoint the important characteristics.

• **Comprehensibility**. Explanations must be comprehensible to humans. This belongs on the target audience and has psychological and social implications, although short explanations ordinarily go a long way across comprehensibility.

Miller studied explainability from the social sciences perspective [110] and notes four essential observations: (i) people prioritize contrastive explanations, i.e. why the model took a specific decision does not matter to us as much as why a different decision was not taken instead; (ii) people choose only a few reasons from the various reasons that make up an explanation, and personal biases evidence this selection; (iii) refer ring to probabilities orstatistical links is not as efficient as refer ring to reasons; and (iv) explanations are social, and thus should be portion of a larger conversation, or an interaction between the explainer and the explainee.

In [58], the authors confirm the importance of human domain experts guiding the growth and evaluation of explanation paradigms, given that current machine learning frameworks work on a statistical and/or model-free setting, and demand context from

human/scientific frameworks to transfer convincing explanations (particularly for other domain experts). No single explanation model in the current literature works is able to satisfy all the mentioned properties (for more details refer to [16, 52, 112] for extensive surveys workbooks on explainable artificial intelligence frameworks).

In our frameworks, all these points were taken into consideration. Furthermore, as mentioned in Chapters 3, 4, 5 and 6, our contributions/publications achieved all of these points.

1.6 Repercussion the limits of biometric on kinship systems

Kinship verification systems suffer and affected by traditional factors of biometric systems as well. Some of the main factors affecting the accuracy of the biometric systems [71] are as follows:

1. **Noise in sensed data:** Noise in the obtained biometric face sample could result from degraded and improperly maintained cameras or unfavorable ambient (unconstrained environment) conditions. For instance, quality camera also could result in a noisy face image as shown in Fig. 1.6. Noisy biometric face sample could not be felicitous matched, for genuine users, by their competent templates in the dataset or cloud infelicitous matched with the impostors, in which leading to a considerable minimisation in the performance of the framework [160, 161].

Figure 1.6: Typical frames from surveillance videos. (a) and (c) are the surveillance images from a camera with CIF size (pixels) and a camera with 720P size (pixels) respectively; (b) shows two noisy interested faces extracted from (a) and (c) [73].

2. **Intra-class variations:** Intra-class variations in biometric face samples are exemplary created by the user's unsuitable interaction with the time span at capture or the camera e.g., incorrect facial pose - see Fig. 1.7), changes in the environment conditions (e.g., illumination changes) [182], use of various cameras during enrollment and verification, or temporal-related variation in the biometric features such as aging [36]. Huge intra-class variations generally reduce the genuine acceptance rate (GAR) of a biometric system.

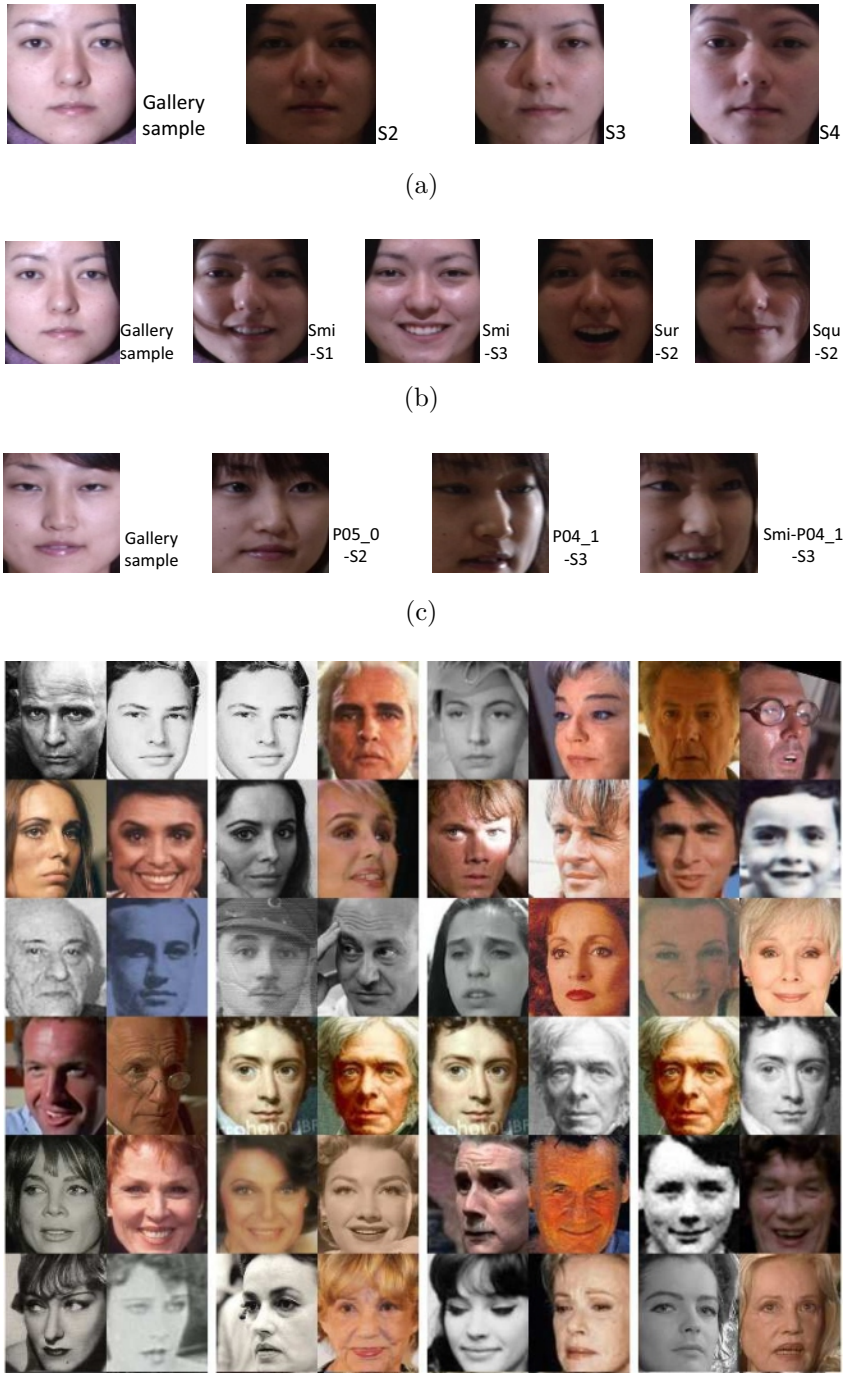


Figure 1.7: Example face images with (a) illumination variations in different sessions [182], (b) expression variations in different sessions [182], (c) pose variations in different sessions [182], and (d) Positive sample pairs from AgeDB [113] with the gap of 30 years, facial appearances undergo dramatically changes in this time span [36].

3. **Inter-class similarities:** Inter-class similarity is known as the interference of the biometric samples, in the feature level space, according to different classes or peoples. The weak of singularity in the biometric trait set leading to a maximize in the false acceptance rate (FAR) of the framework. Therefore, there is a maximum bound on the number of singular individuals that could be assimilation by the biometric framework.

4. **Non-universality:** Universality denotes that each person utilizing a biometric framework is capable of ready the respective biometric feature. The biometric framework could not be capable to extract significative biometric information from a dataset of users. For example, the National Institute of Standards and Technology (NIST) has reported that it is impossible to extract correct minutia traits from the fingerprints of two persons of the population (manual workers with various bruises and cuts on their own fingertips, individuals with hand-related inabilities etc.), due to the low quality of the ridges [21]. This contributes to maximize in the failure to enroll (FTE) rate. Therefore, there is no biometric feature is truly universal.

5. **Interoperability issues:** Generality biometric frameworks are contagious and designed by the presumption that the biometric sample to be verified is acquired utilizing the same camera and, therefore, are restricted in their capability to match or verify biometric samples resulting from various cameras.

6. **Spoof attacks:** A biometric spoof attack is the intentional attempt to tamper one's biometric features in order to dodge verification, or the induction of physical biometric artifacts in order to reincarnate on the identity of another individual.

In kinship verification field, all the proposed methods must be face all these challenges (challenges of biometric systems) before dealing with the kinship factor. In Chapters 3,4,5 and 6, our results performed by the proposed frameworks show that all these challenges are treated successfully over kinship verification databases captured in unconstrained environments.

1.7 Performance evaluation

Biometrics is the scientific term for body measurements and computations. It indicates to metrics belonged to human traits. Biometrics verification (or realistic authentication) is utilized in computer vision as a form of matching identity and accessing control. We refer that also utilized to identify persons in groups that are under surveillance.

Biometric recognizers are the discriminatory, measurable traits utilized to describe and label persons. Biometric recognizers are overwhelmingly classified as physiological against behavioral traits. Physiological traits are belonged to the the body shape. Examples involve, but are not limited to palm veins, fingerprint, facial recognition, palm print, DNA, hand geometry, retina and odour/scent, iris recognition. Behavioral traits are belonged to the modality of behavior of a individual, involving but not limited to typing rhythm, gait, and voice. Some researchers have used the term behaviometrics to characterize the latter class of biometrics.

More conventional means of accessing control comprise token-based recognition systems, such as a driver's license or passport, and knowledge-based identification systems, such

as a password or personal identification number. Since biometric identifiers are singular to each person, but they are highly reliable in the identity verification than token and knowledge-based methods; furthermore, the set of biometric identifiers elevates privacy issues about the integral utilize of this information traits.

1.7.1 Biometric functionality

Many various aspects of person physiology, behavior or chemistry may be utilized for biometric verification. The choose of a specific biometric for utilize in a particular application includes a weighting of various factors. Jain et al. [69] describe seven important factors (indication points) to be utilized when estimating the suitability of any feature for utilize in biometric verification.

- Universality means that each person utilizing a system must take possession of the feature.
- Uniqueness means the trait must be sufficiently varied for persons in the pertinent population such that they may be differentiated from one another.
- Permanence belongs to the way in which a feature changes over time. More particularly, a feature with 'good' permanence will be rationally unchanged over time with conserve to the particular matching model.
- Measurability (collectability) belongs to the facility of measurement and/or storage of the feature. Also, acquired information must be in a shape that simply permits subsequent treatment and extraction of the pertinent trait sets.
- Performance belongs to the speed, accuracy, and robustness and effectiveness of technology utilized (see performance subsection 1.7.2 for more specifics).
- Acceptability belongs to how highly persons in the pertinent population consent the technology in which that they are ready to have their biometric feature captured and processed.
- Circumvention belongs to the simply with which a feature should be imitated utilizing a substitute or an artifact.

Proper biometric utilize is extremely application dependent. Furthermore, certain biometrics should be better than other ones based on the in demand levels of security convenience [17]. No monocular biometric should meet all the requirements of all possible application [69].

The block diagram of bimetric system includes two essential modes of a biometric models [70]. First, in verification (or authentication) model the system executes a one-to-one verification of a measured biometric with a special template stocked in a biometric

dataset in which simpler verifying the individual target is the same person they should to be. In general, three procedures are include in the person identity verification [139]. In the first stage, reference frameworks for all users where they are produced and stocked in the framework dataset. In the second stage, some of the samples where they are matched with reference frameworks to produce the impostor and genuine scores and compute the threshold. The third stage is the testing stage. This step may utilize a smart card (SC), ID number or username (e.g. PIN) to signalize which template must be utilized for verification comparison. 'Positive identification' is a joint use of the authentication mode, "where the goal is to block multiple users from utilizing the unique identity" [139].

Second, the identification/recognition mode, the framework proceed a one-to-many verification/comparison versus a biometric dataset in an effort to found the identity of an unknown person. The system could succeed to manage the identifying of the person if the verification of the biometric sample tested to a template in the datatset falls within a previously set threshold. Recognition/Identification mode may be utilized many for 'true recognition' (so that the user doesn't have to show any information data about the template to be utilized) or for 'false recognition' of the individual "where the framework founds whether the individual is who she (explicitly or implicitly) denies to be" [139]. This latter may only be done over biometrics since the other approaches of personal identification such as PINs, passwords or keys are unavailing.

1.7.2 Performance

In the following, the utilized as performance metrics for biometric systems (for more details, see Appendix B):

- **False match rate** (FMR, also named as FAR = False Accept Rate): represent the probability of that the framework misclassified the test input pattern to a non-matching sample in the dataset. It represents the percent of null and void inputs that are wrongly accepted. In situation of similarity measure, if the individual is an imposter in reality, but the matched score is larger than the threshold, then we treated it as genuine. This maximizes the FMR, in which thus also relies onto the threshold score [139].

- **False non-match rate** (FNMR, also named FRR = False Reject Rate): the likelihood that the framework wrongly indicate that there is a match between the sample input pattern and a matched template in the dataset. It computes the percent of useful inputs that are wrongly rejected.

- **Receiver operating characteristic** or relative operating characteristic (ROC): The ROC is a visual plot characterization of the trade-off between the FMR (FAR) and the FNMR (FRR). Generally, the matching approach generates a decision based

on a threshold that defines how close to a template the sample input necessarily to be for it to be look as a match. When the threshold is decreased, there should be smallest false non-matches but additional false accepts. Furthermore, a larger threshold should decrease the FMR but enlarge the FNMR. A common difference is the Detection error trade-off (DET), by which it is acquired utilizing normal of deviation scales on the two axes. This more linear graph lighten the divergences for greater performances (scarce errors).

- **Equal error rate** or crossover error rate (EER or CER): the rate at which both rejection and acceptance errors are equal. The rate of the EER can be simply extracted using the ROC curve. The EER is a rapid way to be compare the accuracy of machines with various ROC curves. Generally, the machine with the smallest EER is the high accurate.

- **Failure to enroll rate** (FTE or FER): the rate at which tries to generate a template from an sample input is failing. This is high commonly inspired by low-quality sample inputs.

- **Failure to capture rate** (FTC): Within automatic frameworks, the likelihood that the framework fails to determine a biometric sample input when given correctly.

- **Template capacity**: the extreme number of collections of data that may be stocked in the framework.

1.8 Articulation of the thesis

The thesis manuscript is structured around seven chapters:

In Chapter 1, we gave a general introduction of the contexts, motivations, objectives and contributions of this thesis.

In Chapter 2, we mention a general overview state of the art methods of the kinship verification as well as their different types: features learning-based kinship verification, metric learning-based kinship verification and convolutional deep learning-based kinship verification. On the other hand, we presented the kinship problem from facial images and its measuring characteristics as well as the general kinship verification system.

In Chapter 3 (our first contribution), we present a Facial Kinship Verification (FKV) approach based on an automatic and more efficient two-step learning into color/texture information. Most of the proposed methods in automatic kinship verification from face images consider the luminance information only (i.e. gray-scale) and exclude the chrominance information (i.e. color) that can be helpful, as an additional cue, for predicting relationships. We explore the joint use of color-texture information from the chrominance and the luminance channels by extracting complementary low-level features from different

color spaces. More specifically, the features are extracted from each color channel of the face image and fused to achieve better discrimination. We investigate different descriptors on the existing face kinship databases, illustrating the usefulness of color information, compared with the gray-scale counterparts, in seven various color spaces. Especially, we generate from each color space three subspaces projection matrices and then score fusion methodology to fuse three distances belonging to each test pair face images. Experiments on three benchmark databases, namely the Cornell KinFace, the KinFaceW (I & II) and the TSKinFace database, show superior results compared to the state of the art.

In Chapter 4 (our second contribution), we present the combination of deep and shallow features (multi-view features) using the proposed metric learning (SILD+WCCN/LR) approach for kinship verification. Our approach based on an automatic and more efficient two-step learning into deep/shallow information. First, five layers for deep features and five shallow features (i.e. texture and shape), representing more precisely facial features involved in kinship relations (Father-Son, Father-Daughter, Mother-Son, and Mother-Daughter) are used to train the proposed Side-Information based Linear Discriminant Analysis integrating Within Class Covariance Normalization (SILD+WCCN) method. Then, each of the features projected through the discriminative subspace of the proposed SILD+WCCN metric learning method. Finally, a Logistic Regression (LR) method is used to fuse the six scores of the projected features. To show the effectiveness of our SILD+WCNN method, we do some experiments on LFW database. In term of evaluation, the proposed automatic Facial Kinship Verification (FKV) is compared with existing ones to show its effectiveness, using two challenging kinship databases. The experimental results showed the superiority of our FKV against existing ones for bi-subject matching on KinFaceW-II and TSKinFace. Also, the experimental results showed the superiority of our FKV on the available TSKinFace database for Father-Mother-Son and Father-Mother-Daughter.

In Chapter 5 (our third contribution), we develop a novel criterion, named Side-Information based Weighted Exponential Discriminant Analysis (SIWEDA), that is based on the classical SIEDA method. We reformulate and generalize the classical Fisher criterion function in order to maximize it, with the property to pull as close as possible the intra-class samples (within-class samples), and push and repulse away as far as possible the inter-class samples (between-class samples). Thus, SIWEDA selects the eigenvalues of high significance and eliminate those with less discriminative information. To reduce the feature vector dimensionality and lighten the class intra-variability, we use SIWEDA and within class covariance normalization (WCCN) using the proposed statistical binarized image features (StatBIF). Moreover, we use score fusion strategy to extract the complementarity of different weighting scales of our StatBIF descriptor. We conducted experiments to evaluate the performance of the proposed method under unconstrained environment, using five datasets namely LFW, YTF, Cornell KinFace, UB KinFace and TSKinFace datasets,

in the context of matching faces and kinship verification in the wild conditions. The experiments showed that the proposed approach outperforms the current state of the art. Very interestingly, our approach showed superior performance compared to methods based on deep metric learning.

In Chapter 6 (our fourth contribution), we present a new Tensor Cross-view Quadratic Discriminant Analysis (TXQDA) method based on the XQDA method for kinship verification in the wild. Many researchers used metric learning methods and have achieved reasonably good performance in kinship verification, none of these methods looks at the kinship verification as a cross-view matching problem. To tackle this issue, we propose a tensor cross-view method to train multilinear data using local histograms of local features descriptors. Therefore, we learn a hierarchical tensor transformation to project each pair face images into the same implicit feature space, in which the distance of each positive pair is minimized and that of each negative pair is maximized. Moreover, TXQDA was proposed to separate the multifactor structure of face images (i.e. kinship, age, gender, expression, illumination and pose) from different dimensions of the tensor. Thus, our TXQDA achieves better classification results through discovering a low dimensional tensor subspace that enlarges the margin of different kin relation classes. Experimental evaluation on five challenging databases namely Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases, show that the proposed TXQDA significantly outperforms the current state of the art. In addition, our TXQDA method works well on smallest or limited training data classes and on biggest or large-scale training data classes.

In Chapter 7, we conclude this thesis by summarizing the main points of our contributions and we mention some interesting perspectives to be explored following our work.

2 State of the art review

Contents

2.1 Introduction

Kinship verification models consists in checking if two persons are belonging to the same family or not is termed kinship (or family) verification. Automatic kinship verification aiming to discover computational models to decide whether two persons are from the same family or not and purely based on patterns such as voices, faces and gaits. The automatic kinship verification systems define their inputs by two faces (Face A and Face B) and the predictable output is a decision whether Person A is with relation with a family member (father, mother, brother, sister etc.) of Person B or not. Many applications can be beneficial e.g. for forensics, finding missing children, social media comprehension and image annotation. Though a DNA test is the most precise way for kinship verification, it regrettably cannot be used in many situations such as in video surveillance.

The existing works on kinship verification essentially share similar face features as in face recognition. This involves for instance the use of shallow features LBP (Local Binary Patterns), LPQ (Local Phase Quantization) and HOG (Histograms Of Gradients) features for inputs to SVMs (Support Vector Machines) for verification of kin relation [44, 100]. Such methods work better under some limited face image variations (in terms of image resolution, illumination, blur etc.) but always to suffer under unconstrained environment or to generalize to unseen data. However, the very recently developments in machine learning suggest that highest performance can be obtained from learned features e.g. based on deep learning methods [174] instead of shallow features e.g. LBP, LPQ and HOG.

Many authors feed their method by different features or multiple features (multi-view data) to represent facial images for kinship verification. Lu et al. used the Multiview neighborhood repulsed metric learning (MNRML) [103] method to train four multi-view features, Local Binary Patterns (LBP), Learning-based descriptor (LE), SIFT and Three-patch LBP (TPLBP). Yan et al. [174] employed three different feature descriptors including Local Binary Patterns(LBP), Spatial Pyramid LEarning (SPLE) and Scale-Invariant Feature Transform (SIFT) to extract different and complementary information from each face image through DMML method. Yan et al. [175] applied three dif-ferent feature descriptors including LBP, spatial pyramid lEarning (SPLE), and SIFT to extract different and complementary information from each face image to train the MPDFL method. Lu et al. [101] used four features as it; Local Binary Patterns (LBP), Dense SIFT (DSIFT), the histogram of oriented gradients (HOG) and LPQ for train DDMML method. Lu et al. [60] used MvDML to train four multi-view features, Local Binary Patterns (LBP), Learning-based descriptor (LE), SIFT and Three-patch LBP (TPLBP). Laiadi et al. [83] used three features LPQ, BSIF and CoALBP to train SIEDA method. Dornaika et al. used MNRML to train the two features, FC7 layers of VGG-F and VGG-Face for the purpose of kinship verification. Laiadi et al. proposed TXQDA [84] method to train LPQ and BSIF features using ten scales.

2.2 Kinship verification

Many psychology researches [5, 28, 33, 74, 75] studied the human perception of kinship verification aiming to understand the ability of kin inference from faces. Motivated by this research, machine learning and computer vision communities are showing increasing interest in incubating and promoting computational approaches to verify kin relations between humans. Although several works have been published in the recent years, satisfactory results are still beyond. To the best of our knowledge, the work of Fang *et al.* [44] was among the first effort to deal with the challenge of kinship verification from face images by collecting the first database containing kin-related image pairs. They utilized 22 features (such as distance from eye to nose, skin color, etc.) for kinship classification. Firstly, by using a simplified pictorial structure model, they localized the essential facial features in an image, which are extracted to characterize the face. Then, they computed the differences between feature vectors of corresponding parent and child, and applied the k-nearest-neighbor (KNN) and support vector machine (SVM) classifiers to classify the pairs of face images. More recently, Xia *et al.* [169, 170] proposed a transfer subspace learning algorithm for kinship verification. Their main idea is to use an intermediate young parent set of face images to minimize the dissimilarity between the children and old parent images, basing on the hypothesis that the children and young parents have more facial similarity.

Later on, Lu *et al.* [103] released two databases, KinFaceI & KinFaceII, for kinship verification which greatly promoted the research on the topic. They also suggested a neighborhood repulsed metric learning (NRML) method for kinship verification. The purpose is to learn a distance metric with the property to pull as close as possible the intra-class samples (with a kinship relation), push and repulse away as far as possible the inter-class samples (without a kinship relation) lying in a neighborhood. Similarly, Yan *et al.* [62, 174], proposed a discriminative multimetric learning method for kinship verification through facial image analysis. Firstly, they extracted various features to describe face images from several aspects in order to obtain integral information. Then, the multimetric approach was applied to different features to learn suitable metrics for each one.

Dehghan *et al.* [35] employed autoencoders with a discriminative neural network layer to learn both the features and metrics. They proposed an algorithm that integrates the two techniques to determinate parent-offspring relationships. Moreover, they examined and discussed the interconnection between the automatically detected features and those in anthropological studies. Liu *et al.* [96], proposed inheritable Fisher vector feature (IFVF) method. First, the Fisher vector is extracted from each image by assembling the intensity sampled SIFT features from the RGB color space. Second, a new inheritable transformation, which simultaneously increases the similarity between kinship images while decreasing that between non-kinship images, is learned based on the Fisher vectors. As a

result, and from each image, the IFVF is extracted by using the inheritable transformation applied on the Fisher vector. The authors also applied a fractional power cosine similarity measure for kinship verification. Recently, Wu *et al.* [168] investigated for the first time the usefulness of color information in the verification of kinship from facial images by using a simple scoring approach. For this purpose, they extracted joint color-texture features from RGB, HSV and YCbCr spaces using different descriptors. The features are then concatenated to form an enhanced feature vector. Finally, they applied cosine similarity between the feature vectors of the pair of the two face images.

Yan H [172] presented a neighborhood repulsed correlation metric learning (NRCML) method for kinship verification through facial image analysis. The author utilized the correlation similarity measure where the kin relation of facial images can be better highlighted. Since negative kinship samples are usually less correlated than positive samples, the most discriminative negative samples are automatically identified in the training set to learn the distance metric so that the most discriminative information encoded by negative samples can be better exploited.

In summary, the existing facial kinship verification methods could be categorized into three groups: (a) feature learning-based [2, 20, 38, 44, 49, 86, 87, 98, 105, 114, 156, 168, 170, 173, 175]. (b) metric learning-based [14, 39, 61, 101, 103, 131, 193, 195]. (c) convolutional deep learning-based [79, 80, 155, 157, 177, 178, 188, 189].

2.3 Measuring kinship characteristics

Searching a true facial representation is the solution to any face analysis framework. If the obtained facial traits are really effectively discriminative, one could easy publish the note book as the nearest neighbor. Furthermore, verifying the relationships of kinship, the general idea of this kind of research was to used certain discriminative traits extracted from facial images of cut facials to obtain stable cues linked to kinship. In the following, a review of the several features and traits that have been developed and proposed in the literature works to verify kinship.

2.4 The general kinship verification framework

Due to the hardness and difficulty of kinship verification task, we look at the various differences at several stages with the division strategy to conquer them, as shown in Fig. 2.1. We can see that the architecture scheme is roughly subdivided into four components, namely preprocessing, feature extraction and the degree of similarity to the kinship and verification [132]. It is necessary to note that we only provide the general tasks for the facial kinship verification task and that not all blocks are important for instantiation.

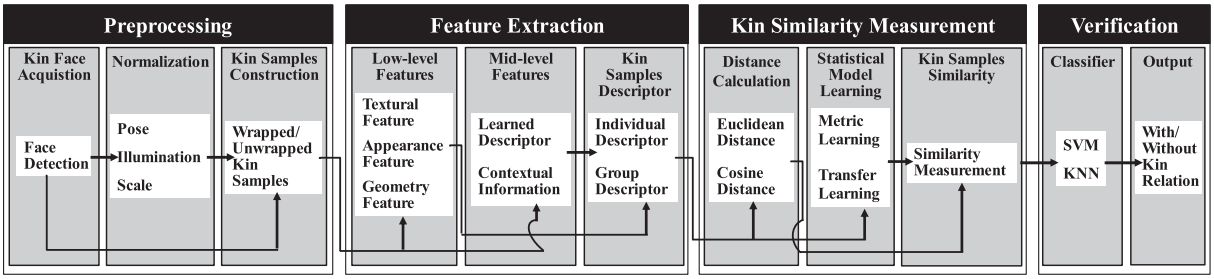


Figure 2.1: A global system framework for kinship verification [132].

2.5 Features learning-based kinship verification

In this section, we cite and illustrate the state of the art methods which learn features/traits of kinship to describe the face images.

Fang et al. [44] considered the first one who propose kinship verification over facial images, where they used a set of low-level features. They make a test and evaluation of the individual performance of numerous low-level traits of face images, and then select the best 14 features. We split these best features into three parts: the color, face parts distances, and the gradient histograms. More recently, Lopez et al. [98] proposed to make prediction over the use of the chrominance distance metric between each pair of face images as the confidence score. These types of approaches can extend to a certain precision, but there is the existence of many problems, amongst that the hypothesis of each face kinship pair is cropped using the same photo and the pose variation of face images is comparatively simple are generally outstanding problems.

Xia *et al.* [170] provided two model features that simply can separate positive child–parent face images pairs from the negative ones. The first one was based on the facial appearance and taken by 40 collection of Gabor [1] filters parameters (eight directions and five scales). MoreoverFurthermore, they firstly subdivided each of the facial image into regions in five parts, as shown in Fig. 2.2. As shown in this figure, the total face image is the first part. Furthermore, the second part includes many regions as it: upper, lower, left, right and center regions of the face image. The brow, nose, eyes, mouth, and the areas of cheek are include the third part and their finer sub-parts form the fourth part. Finally, a collection of sub parts based on the four fiducial points form the fifth part. After that, they used Gabor filters on each aforementioned local parts. The same idea has been utilized in [164] to describe the face expressions through local traits description. Intuitively, facial kinship verification is also a process on local parts. Meaning that when people are talking about face kinship, they predominantly focus on parts of the face between parents and their children and decide whether they share comparable eyes, noses, or mouths. Another trait is based on the anthropometric model [135] which basically looks for structure data of facials. Based on the captured key points, they obtained 6-D structure traits of ratios of typical region distances, e.g., “eye–eye” metric versus “eye–nose” metric. Structured data

is supposed to inherit widely from parents, and therefore might be a key step for kinship verification [44]. However, due to aging factor between the parents and their children [135], the old parents facial structures are transformed from the ones when they were young. So, they use transfer learning subspace to lighten the age degrading factor.

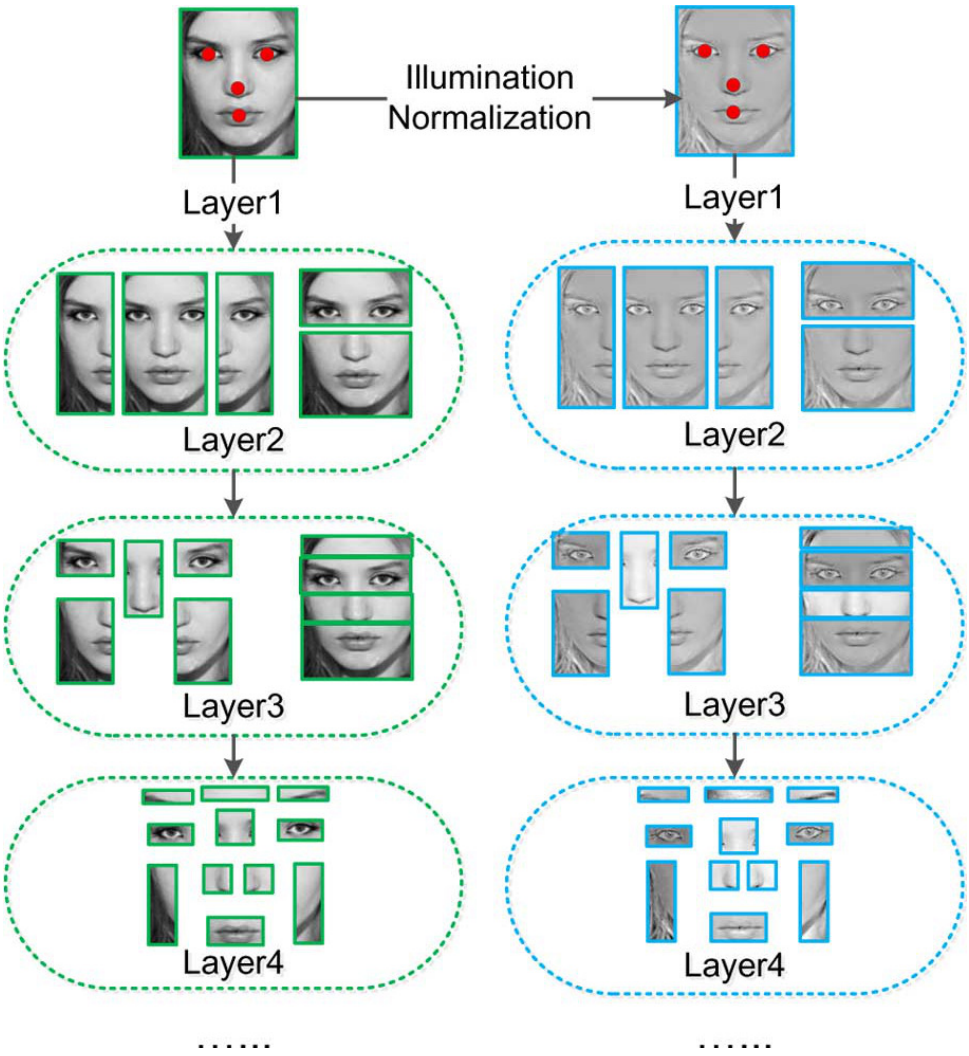


Figure 2.2: Face partitions in different layers and face image illumination normalization. For simplicity, only four layers are illustrated instead of five. Red dots on faces in Layer1 illustrate four key points mentioned in this work [170].

As seen in Fig. 2.3, Wang et al. [170] utilised the height differences and the closeness of persons change in family and non-family face images. Under a reasonable hypothesis of camera pose. The camera positions of faces in an image supply an estimate of proportional height of people in the image (which may signalize their proportional age), and proportional physical nearness in the photograph (which may signalize social context). Although these estimates may not be true, the total geometry of face positions is a key step for family photo classification. The geometric data of a group facial images has shown to be more helpful for event recognition [46]. For social relationship analysis in a group face image, there have been efforts to add this new factor into account. Pairwise measure of two persons' positions is the most generally utilized method. Wang et al. [153] utilized the metric

between two individuals to represent the closeness in a face image. Counting the number of individuals between the samples being compared was utilized in [25]. Furthermore, these measurements only measure the relative proportional of two people, instead of the total geometry as we longing.

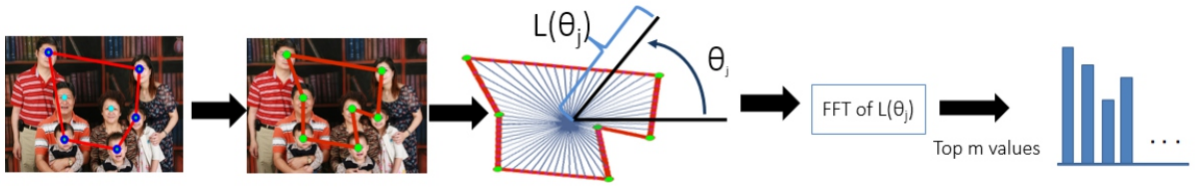


Figure 2.3: Face partitions in different layers and face image illumination normalization. For simplicity, only four layers are illustrated instead of five. Red dots on faces in Layer1 illustrate four key points [170].

Bottinok et al. [20] relies upon more concentrate on texture features, including LPQ, WLD, TPLBP and FPLBP. The proposed framework products the highest performance on the image restrict setting in the second kinship competition [100]. Furthermore, Bottino et al. [19] and Vieira et al. [151] used a collection of geometric, holistic or texture traits. These two approaches got 81.5% on KinFaceW-II database. The results show that multiple traits instead of monocular trait is more powerful to got a higher performance for kinship verification.

Besides the geometry, color and texture traits, appearance features is another efficient path to characterize kin facials due to the gaining obtained in facial analysis task. This kind of method comprises Gabor wavelet [196], salient part [54] and self similarity [78].

Wang et al. [156] used both features, the appearance and geometry traits. For the appearance trait, pyramid facial images are first constructed on each facial to get an overlapping blocks then feature extraction is performs within those blocks. A Gaussian blend model is used to find the similar block pairs in corresponding locations of two facial images. Then, the absolute gap between two similar blocks traits is calculated as the appearance trait. For geometry trait, face landmarks are first calculated and then projected to a new subspace called the Grassmann manifold. Finally, the Geodesic metric between two facial frames is computed as the geometry trait.

Moreover, Dibeklioglu et al. [38] show that facial dynamics may provide more effective information cues for facial kinship verification. To this end, videos of expressions (enjoyment smiles) instead of still face images are utilized to extract facial dynamic traits fused with spatio-temporal appearance description. The prudence of this approach is that the casual facial expressions of born-blind individual and their considered family are similar [129] not only depends on the appearance of the expression but also related to its dynamics.

More recently, many of researches have been performed on feature learning in the domain of computer vision, and a lot of different types of feature learning methods have been presented. In facial kinship verification, representative feature learning methods comprise spatial pyramid learning-based (SPLE) approach [194], gated auto-encoder [35]

and Convolutional Neural Networks [79, 80, 155, 157, 177, 178, 188, 189]. These methods usually learn representations by encoding some prior understanding, such as sparsity, smoothness, or spatial and temporal coherence, directly through raw pixels [13].

Differently from the aforementioned approaches, Yan et al. [175] used mid-level trait by means of low-level features instead of the raw pixels, where the learned trait vector comprises of various decision values from one SVM (support vector machine) hyperplane. Furthermore, a big unlabeled facial dataset and a very small dataset of facial pairs labeled with kinship pairs relations are utilized to augment an target function so that facials with kin relation are prospective to have identical decision scores from the used hyperplane.

The work of Yan et al. [175] proposed a novel prototype-based discriminative feature learning (PDFL) approach for facial kinship verification. Differently from most prior works on kinship verification which utilize low-level hand-crafted (shallow) descriptors including local binary pattern (LBP) and Gabor traits for facial representation, their work goal is to learn a new and more discriminative mid-level features to perfectly describe the kin relation of facial images for kinship verification. To perform this, they collect a group of facial images with unlabeled kinship relation from the LFW (labeled face in the wild) dataset as the baseline set. After that, each face in the training facial kinship database is demonstrated as a mid-level feature (trait) vector, where each input is the identical decision score from one SVM hyperplane. Thereafter, they design an target function by decreasing the intra-class faces (with a kin relation) and increasing the neighboring inter-class faces (without a kin relation) with the mid-level traits. Finally, they used multiple low-level traits for mid-level feature learning. Therfore, they further suggested a multi-view PDFL (MPDFL) approach to learn multiple mid-level traits to improve the verification of kinship performance. Fig. 2.4 shows the illustration of the proposed pipeline for facial kinship verification.

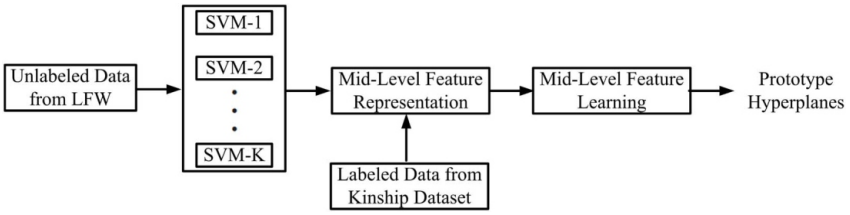


Figure 2.4: Pipeline of the proposed kinship verification approach. First, they construct a set of face samples from the LFW dataset as the prototypes and represent each face image from the kinship dataset as a combination of these prototypes in the hyperplane space. Then, they use the labeled kinship information and learn mid-level features in the hyperplane space to extract more semantic information for feature representation. Lastly, the learned hyperplane parameters are used to represent face images in both the training and testing sets as a discriminative mid-level feature for kinship verification [175].

The work of Moujahid et al. [114] presented a new approach for image-based kinship verification eligible to effectively fuse local and global facial traits information extracted from various descriptors. As shown in Fig. 2.5 the proposed framework depends on two main steps: (1) they model the facial images utilizing a Pyramid Multi-level (PML) description where local face descriptors are extracted through many blocks at various

size scales; (2) they calculated the covariance (second-order statistics) of various local features describing each individual block in the PML description. This allows rise to a facial descriptor with a two effective properties: (i) that the PML description, different scales and facial regions are explicitly fused together in the final description without the need to detect the face landmarks; (ii) the covariance descriptor describes spatial traits of any kind granting the combination of various state-of-the-art color and texture features.

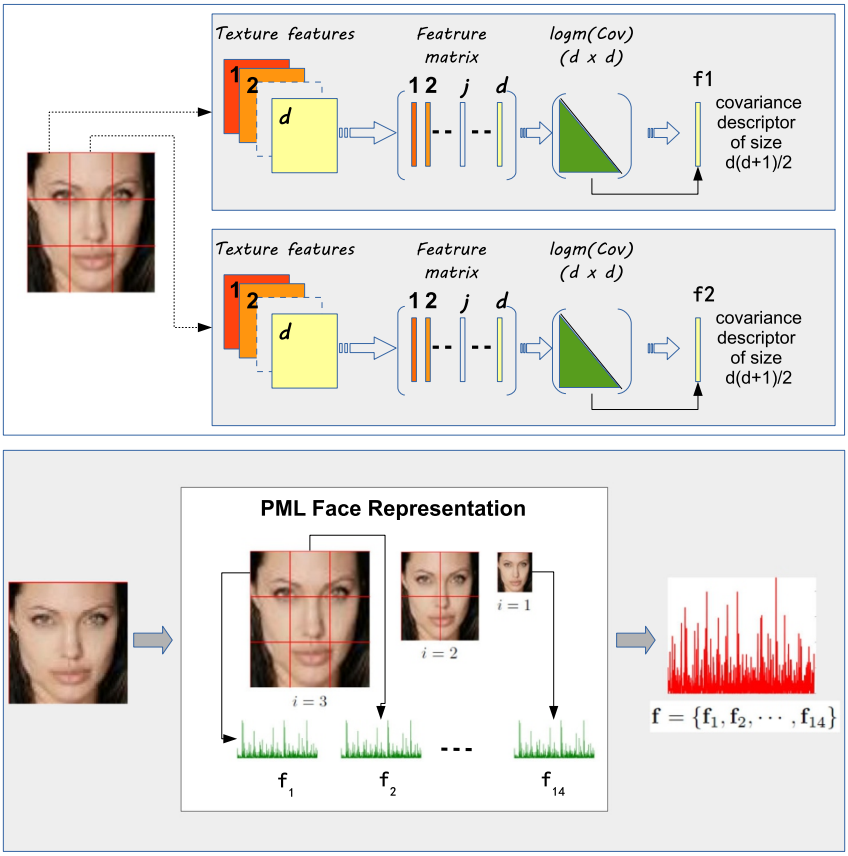


Figure 2.5: (Top) Multi-block covariance descriptor and (Bottom) Pyramid Multi-Level (PML) covariance descriptor [114].

Lan et al. [87] proposed a simpler efficient approach called quaternionic Weber local descriptor (QWLD) for color face image traits extraction. Fig. 2.6 illustrates the proposed QWLD framework for color images. Combining quaternionic representation (QR) of the color facial image and Weber's law (WL), QWLD has both their proprieties. It utilizes QR form to deal with all color channels of the facial image in a global way while saving their relations (neighborhood from different channels), and applied WL to guarantee that the combined descriptors are more robust and more discriminative. Utilizing the QWLD approach, they further discover the quaternionic-increment-based Weber descriptor and quaternionic-distance-based Weber descriptor in terms of multiple perspectives.

Wu et al. [168] realized that the generality of the proposed approaches for facial kinship verification have essentially based on processing only the luminance (i.e.gray-scale) of the facial images, hence excluding the chrominance (i.e. color) information data which may be a powerful additional trait to verify kinship from faces. Their work shows for the first

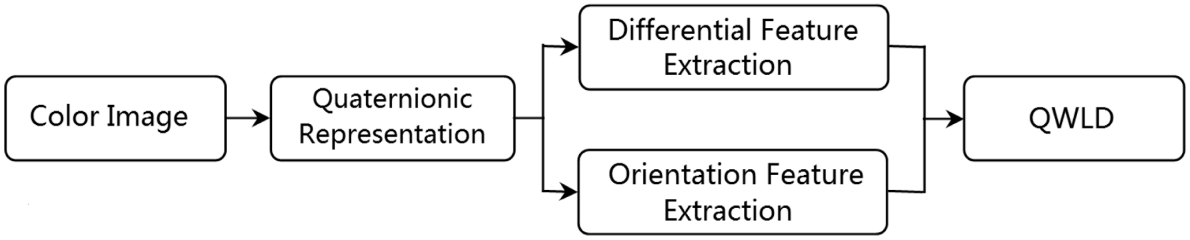


Figure 2.6: QWLD framework [87].

time that the color information contains an additional traits in the verification of kinship from face images. For this objective, they used joint color-texture traits to describe both the luminance and the chrominance data information in the color facial images. The kinship verification performance utilizing both color and texture analysis is compared to the counterpart methods using only gray-scale data information. Fig. 2.7 shows an illustration of the Color/Texture classification method for facial kinship verification.

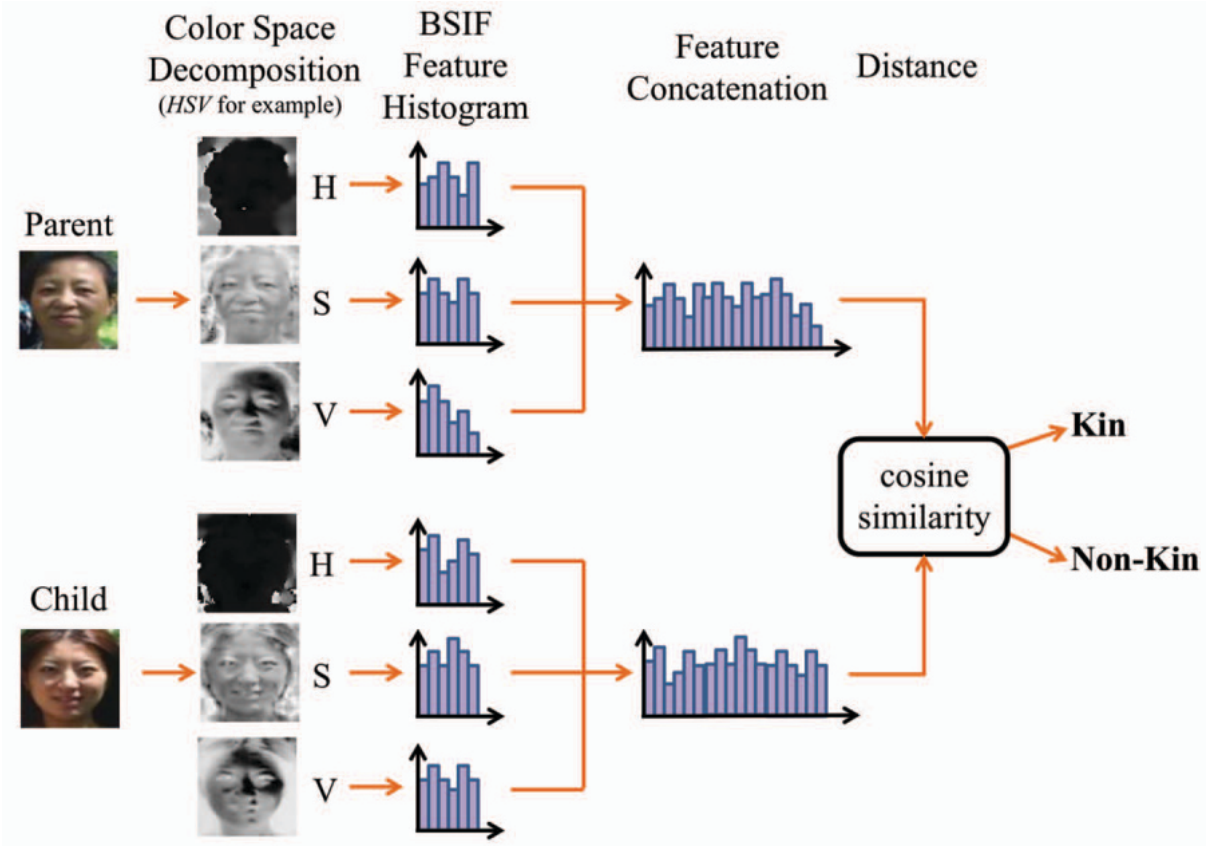


Figure 2.7: An illustration of the Color/Texture classification method [168].

Mahpod et al. [105] proposed a multiview hybrid combined symmetric and asymmetric distance learning (CSADL) network for face kinship verification. Both discriminative descriptions are combined for the parents and the children utilizing a margin maximization learning framework, while the kinship verification is formed as a classification task solved by SVM. Fig. 2.8 describes the proposed hybrid distance learning (HDL) network for

facial kinship verification.

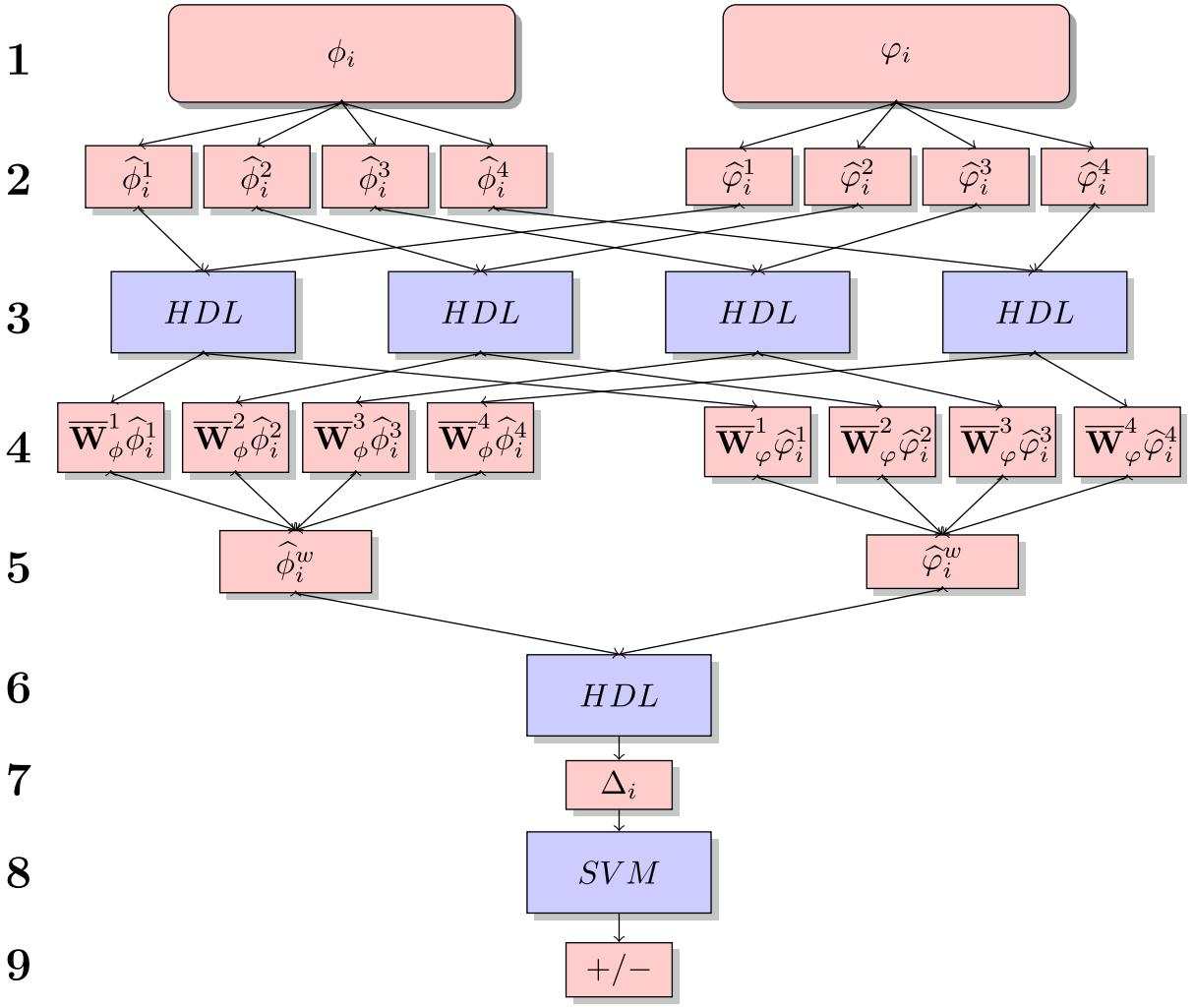


Figure 2.8: The proposed hybrid distance learning (HDL) network. **(1)** A training set consisting of pairs of faces images of parents and their children. **(2)** Computation of image descriptors. **(3)** Training the HDL per feature. **(4)** Applying the HDL projection. **(5)** concatenating the multiple learnt representations. **(6)** Training the HDL using the fused features. **(7)** Fused representation of the pair of input images. **(8)** Kernel SVM classification. **(9)** Kin verification result [105].

Aliradi et al. [2] proposed a novel framework that used discriminative data information, which is focused on the exponential discriminant analysis (DIEDA) fused with various scale descriptions. The histograms of multiple blocks are assembled together to get a high dimensional vector of features, which demonstrates a specific descriptor of the scale. The projected features based histograms for each region used the cosine similarity distance to minimize the feature data vector dimension. Finally, region scores depending to several descriptors extracted at multiple scales are then fused together and compared by utlizing a classifier. Their work feats effective side information data for face matching and kinship verification in the wild conditions (to decide if the facial image pairs are taken from the same person or not). To tackle this problem, they take samples of the face images with unlabeled kinship facial images from the labeled face in the wild dataset as the baseline set. They created an optimized target function by descreasing the intra-class samples (with

a kin relation) and increasing the interclass samples (without a kinship relation) with the proposed framework. Fig. 2.9 shows the architecture of the dieda face and kinship verification system.

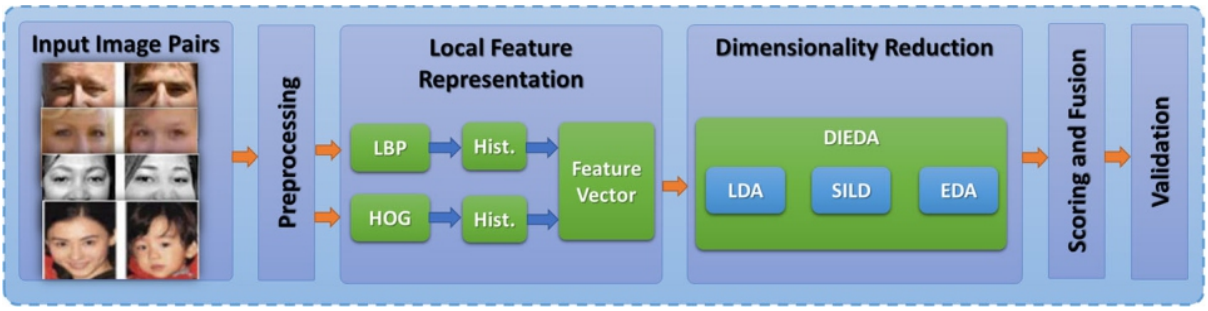


Figure 2.9: The architecture of the dieda face and kinship verification system [2].

Lan et al. [86] suggested a simpler and effective framework named quaternion-Michelson descriptor (QMD) to encode local features for color facial image classification. Fig. 2.10 illustrates schematic picture of Quaternion-Michelson descriptor (QMD). Unlike the most local descriptors using directly from the original (raw face) image data, QMD is deduced from the Michelson contrast law and the quaternionic representation (QR) of color facial images. The Michelson contrast is a robust measurement of facial images tenor from the multiple view paints of human vision, while QR is capable to deal with all the color data information of the facial image holisticly and top save the dynamics over various color channels. Therefore, QMD combines both the merits of Michelson contrast and QR. Based on the QMD approach, the authors further proposed two new quaternionic Michelson contrast binary pattern descriptors from various perspectives.

Figure 2.10: schematic picture of Quaternion-Michelson descriptor (QMD) [86].

Yan et al. [173] proposed a new weakly-supervised (semi-supervised) feature learning approach called discriminative compact binary face descriptor (D-CBFD) for face kinship verification. Unlike the generally previous kinship verification approaches where hand-crafted (shallow features) features are utilized for face description, their D-CBFD performs effective face description from a collect of weakly-labeled data samples. Given a facial image, they first calculated pixel difference vectors (PDVs) at different local regions. Then, they learn a effective projection space to map each PDV and project them into a

new discriminative low-dimensional binary features space, where the total energy data information of the PDV should be well saved and the metric of the positive pairs is smaller and that of the negative pairs is greater. Finally, they pool all binary features vectors over each facial into a histogram extraction features as the final description. Fig. 2.11 illustrates the D-CBFD method proposed for kinship verification.

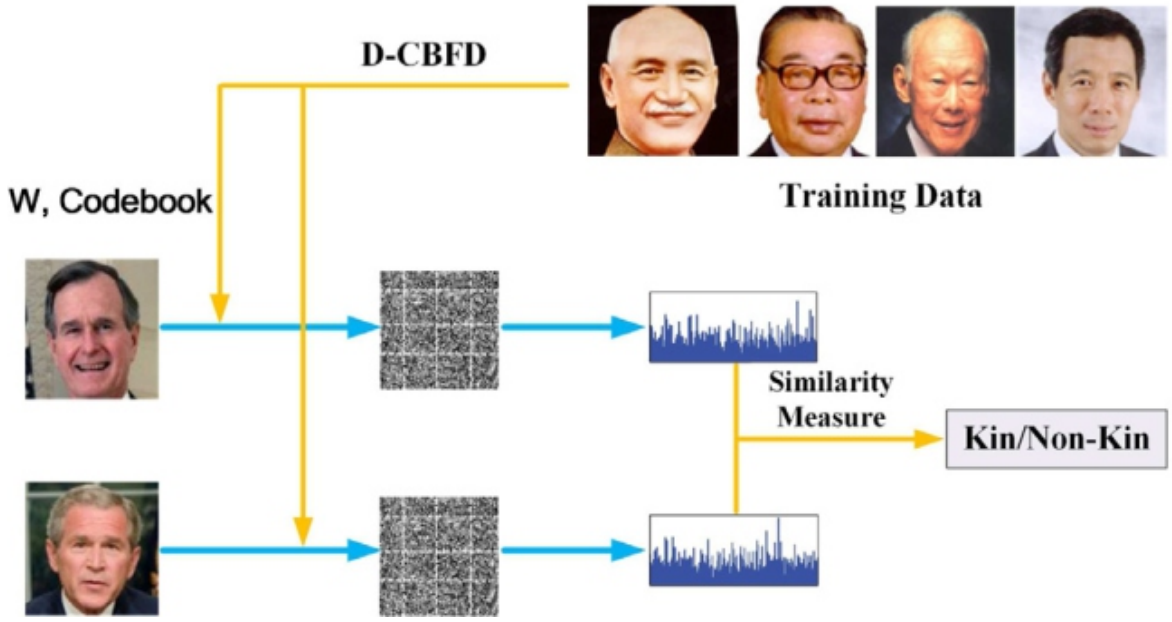


Figure 2.11: The flow-chart of face representation by using D-CBFD. They first divide each training face into several non-overlapped regions and learn the feature mapping **W** and the codebook for each region. Then, they first project each PDV into a low-dimensional binary feature vector. Then, they pool these binary feature vectors within each face into a histogram feature as the final representation [173].

Golay et al. [49] proposed a new eccentricity-based facial kinship verification (EKV) approach to show powerful of dominant face regions for kinship verification. The proposed EKV approach used eccentricity of ellipse-approximated dominant face regions as effective parameter to describe facial images for kinship verification. It shows two essential frameworks, named single eccentricity (SE) and fused eccentricity (FE). SE framework for EKV approach built single formulation by utilizing single face region. For each used face region, it approximated as an ellipse to calculate eccentricity parameter and implement verification. Next, FE framework for EKV approach utilized multi-view description by using two or more face regions. Eccentricity of various ellipse-approximated face regions is calculated and combined to form a converted parameter and implement verification. Fig. 2.12 shows a samples of ellipse estimation for facial images in kinship databases.

Figure 2.12: Samples of ellipse estimation for facial images in kinship databases. The first, second, fifth and sixth columns in each row correspond to original images. The third, fourth, seventh and eighth columns correspond to images with ellipse estimation on facial sections of respective images. Kinship image pairs are distributed as (a) F–S, (b) F–D, (c) M–S, (d) M–D [49].

2.6 Metric learning-based kinship verification

In this section, we show and describe the state of the art methods that learn a metric distance through a deep learning metric learning strategy or subspace transformation metric learning strategy proposed for kinship verification task.

Lu et al. [103] proposed a novel neighborhood repulsed metric learning (NRML) approach for facial kinship verification. Encouraged by the conviction that inter-class samples face images (without a kinship relation) that with huge similarity actually lie in a neighborhood and are more readily misclassified compared to those with less similarity, they goal is to discover a metric distance by which the intra-class samples face images (with a kinship

relation) are pulled as close as possible and inter-class samples face images lying in a neighborhood are repulsed and pushed away as far as possible, Furthermore, such that more effective information can be extracted for kinship verification. To show better utilize of multiple features description and to extract integral information, they further proposed a multiview NRML (MNRML) method to compute a common metric distance to combine multiple features over a subspace fusion to enhance the facial kinship verification performance. Fig. 2.13 shows framework of the proposed kinship verification approach via facial image analysis

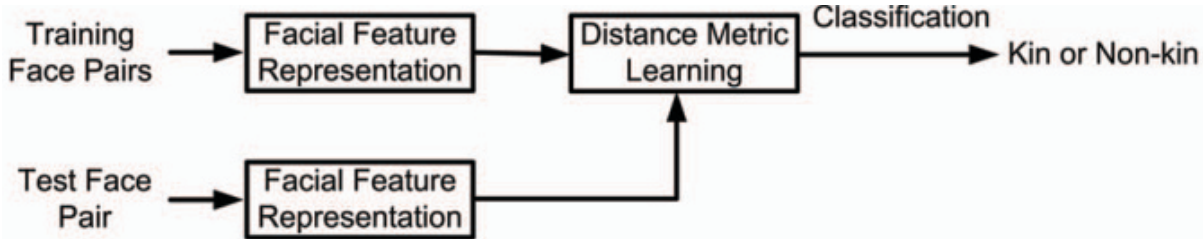


Figure 2.13: Framework of the proposed kinship verification approach via facial image analysis. Given a set of training face images, they first extract features for each face image and learn a distance metric to map these feature representations into a low-dimensional feature subspace, under which the kinship relation of face samples can be better discriminated. For each test face pair, they also extract features of each face image and map these features to the learned low-dimensional feature subspace. Finally, a classifier is used to verify whether there is a kinship relationship or not between the test face pair [103].

Zhou et al. [195] proposed a novel kinship metric learning (KML) approach with a merged deep neural network (DNN) model. As mentioned in Fig. 2.14 KML clearly models the cross-generation contradiction inherent on parent-child pairs face images, and learns a merged deep similarity metric such that the facial image pairs with kinship relation are thrown close (pulled close), while those without kinship relation (but with high appearance similarity) are thrown away (pushed as far away as possible). Furthermore, by assessing the intra-connection assortment and inter-connection uniformity over the merged DNN, they present the property of hierarchical compactness into the merged network to make easier deep metric learning with finite collection of kinship training information data.

The work of Bessaoudi et al. [14] proposed a framework based on tensor-based metric learning (high order tensor) design of facial images. The facial tensor is structured based on local texture descriptors extracted from multi-scales. Furthermore, they proposed a novel Multilinear Side-Information based Discriminant Analysis (MSIDA) to deal the semi-supervised multi-linear subspaces projections reduction and classification. By utilizing only the weakly labeled information data, MSIDA allows to project the input facial tensor into a discriminative subspaces defined by tensor subspaces analysis in which the discrimination is enhanced and the size (dimension) of each tensor modes are minimized simultaneously. As depicted in Fig. 2.15, the block diagram of the MSIDA approach consists of three fundamental components: feature extraction, tensor subspace transformation and comparison.

Each facial image is appeared by two local texture descriptors, MSLPQ and MSBSIF,

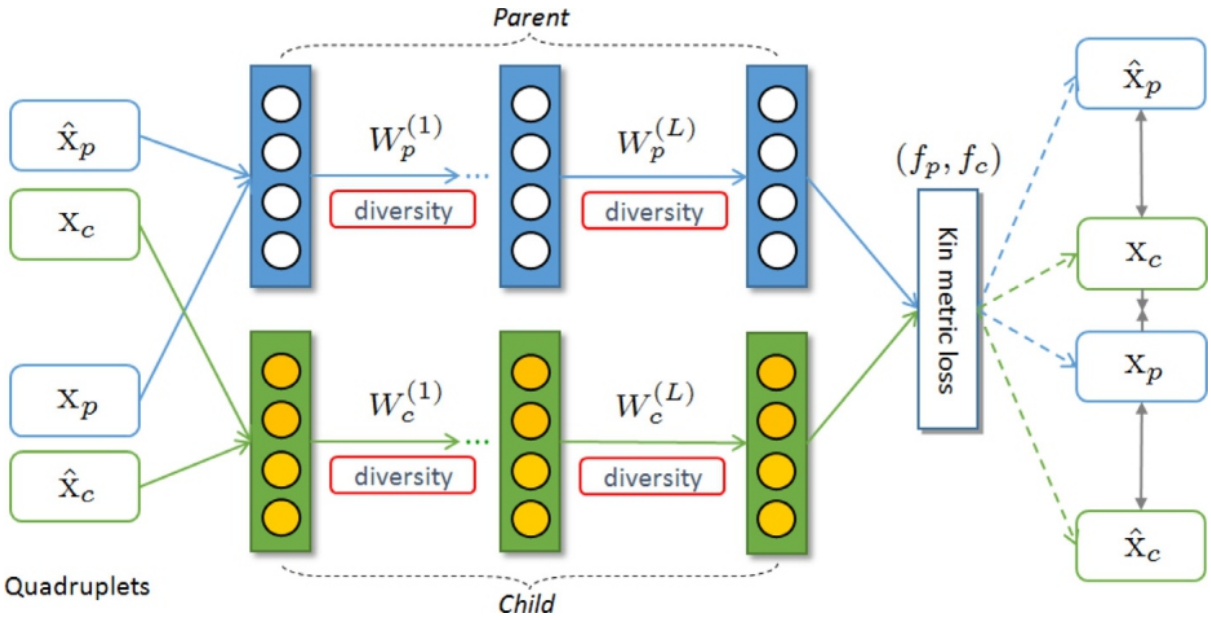


Figure 2.14: The meta-view of the KML method with a carefully designed deep architecture KinNet [195].

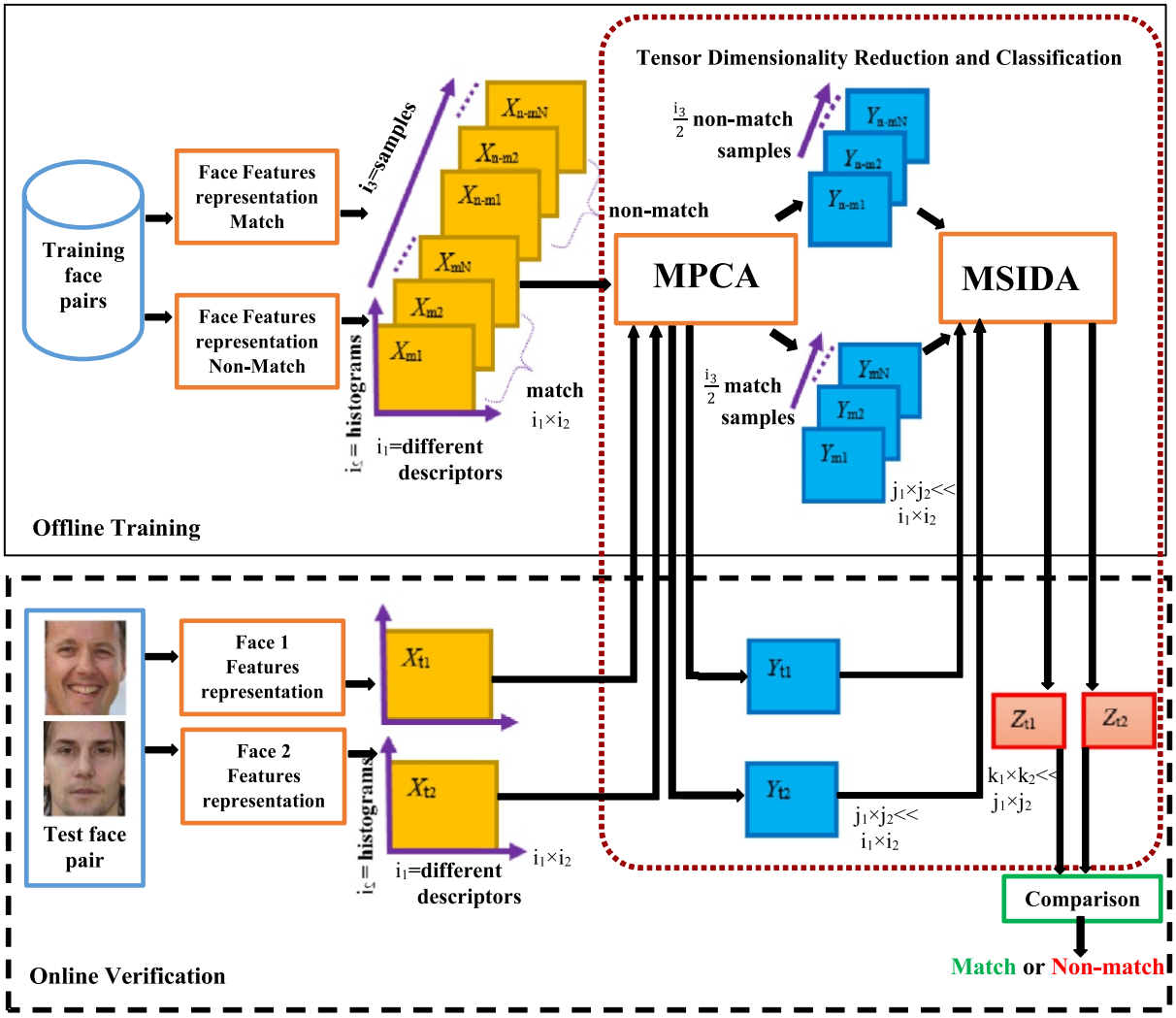


Figure 2.15: Block diagram of the MSIDA face pair matching system [14].

performed at various scales subjecting many features vectors of each face image. The features vectors of the total training facials are stacked as a third order tensor (i_1, i_2, i_3), where i_1 represents the feature length of single feature vector, i_2 represents the various local texture descriptors extracted from multiple scales, and i_3 represents the facial images samples contained in the training dataset. The built tensor is firstly projected by MPCA [154] to minimize the subspace dimensions to $j_1 \times j_2 \times i_3$, where $j_1 \times j_2 << i_2 \times i_2$. The cause for effect MPCA before to MSIDA is to process the small sample size (SSS) problem in several tensor modes. This problem appears when the dimension length of the features vector is bigger than the number of training samples face images, resulting to the singularity of different MSIDA scatter matrices. Minimizing the dimension length of each tensor mode firstly is therefore applied.

After performing MPCA, the training information data tensor was split into two sub-tensors according to the positive pairs (match pairs) and negative pairs (non-match pairs), respectively. The subdivision was done corresponding to the third tensor mode i_3. The positive tensor was utilized to calculate the within-class scatter matrix $(S^{msida})_w$ and the negative tensor was utilized to calculate the between-class scatter matrix $(S^{msida})_b$ of the MSIDA method. The information data tensor was projected through MSIDA subspaces to obtain a lower and more discriminative fearures $k_1 \times k_2$, where $k_1 \times k_2 << j_2 \times j_2$.

In the test stage, each of the facial images of the pair was checked the matching was represented as a second order tensor defined by stacking the local texture descriptors of the facial image. Furthermore, the two tensors were projected though MPCA method and then MSIDA method. Finally, the cosine similarity distance between the test pair was calculated and utilized to decide whether the pair is positive (belonging to the same person/family) or not.

The work of Dornaika et al. [39] introduced a novel scheme that extract facial deep traits for kinship verification. The approach merges effective features selection and kinship-oriented prominent data information projection. The presented framework comprised of three stages of fusion: (1) an early fusion of features descriptors where the filter selection chooses the most discriminant deep features, (2) a middle-level fusion which used a kinship-based multi-view metric learning (MNRML) method, and (3) a late-level fusion that combines classifiers (SVM) responses. In their work, facial features are obtained by the pre-trained deep convolutional neural networks VGG-F and VGG-Face that were basically proposed for classifying groups of objects and identities, respectively.

They focus on four relations of kinship relations. The four relations are: Father–Son (F–S), Father–Daughter (F–D), Mother–Son (M–S), and Mother– Daughter (M–D). Fig. 2.16 shows an overview of the structure of the proposed framework. The input is a pair of facial images. The first image was given to describe a child and the second for the parent. Furthermore, a child facial image according to either a child or an adult. The parent facial image according to a young adult or an old person. The confirm that the order of

positioning of the pairs in the facial images is not important. Therefore, the output of their proposed framework was a binary decision which verifies the facial kinship relation.

Given a pair of facial images (Parent–Child), the deep facial features descriptions by VGG-F and VGG-Face are defined as the vectors **p**, **c** and **p'** and **c'**, respectively. Each was a vector $\in \Re^{4096}$. The processing shown that the data information thrown over a pipeline.

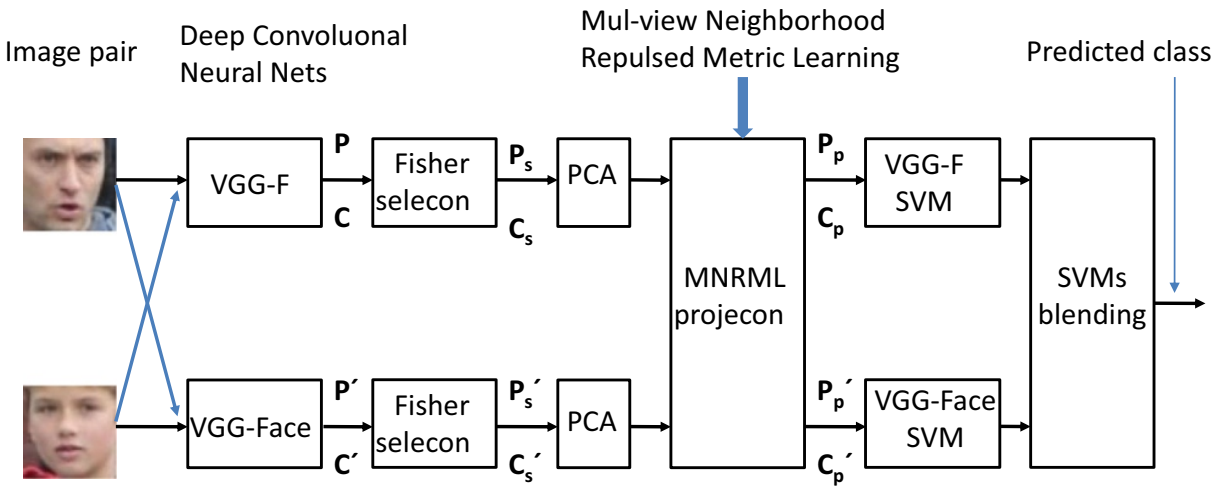


Figure 2.16: Pipeline flowing from left to right, turning the pair of input images into a predicted class. Each block performs a change in its input and its output is connected to the input of another block. The pipeline architecture is the same for any number of descriptor types [39].

Lu et al. [101] proposed a DDML approach to train a deep neural network that can be learn a group of hierarchical non-linear transformations subspaces to project facial images pairs into the same implicit features space, in which the metric of each positive pair was minimized and that of each negative pair was maximized, respectively. To better exploit the commonality of different features descriptors and to make all the features more effective for facial and kinship verification, they developed a efficacious deep multi-metric learning (DDMML) method to together learn multiple neural networks by which the correlation of various features of each sample is increased, and the metric of each positive pair is decreased and that of each negative pair is maximized, respectively.

As illustrated in Fig. 2.17 DDML learns one neural networks from a monocular feature description and cannot deal with various feature descriptions directly. In facial and kinship verification, it is simply to extract various features for each facial image for various feature fusion. However, the features were extracted from the same facial image are generally extremely correlated to each other even if they could describe facial images from several aspects [45]. For various feature fusion, these extremely correlated information data should be saved because they generally reflect the fundamental information data of samples. An significant principle to utilize multi-feature metric learning is to simultaneously learn a multiple distance metrics by saving the correlation between various feature pairs facial images.

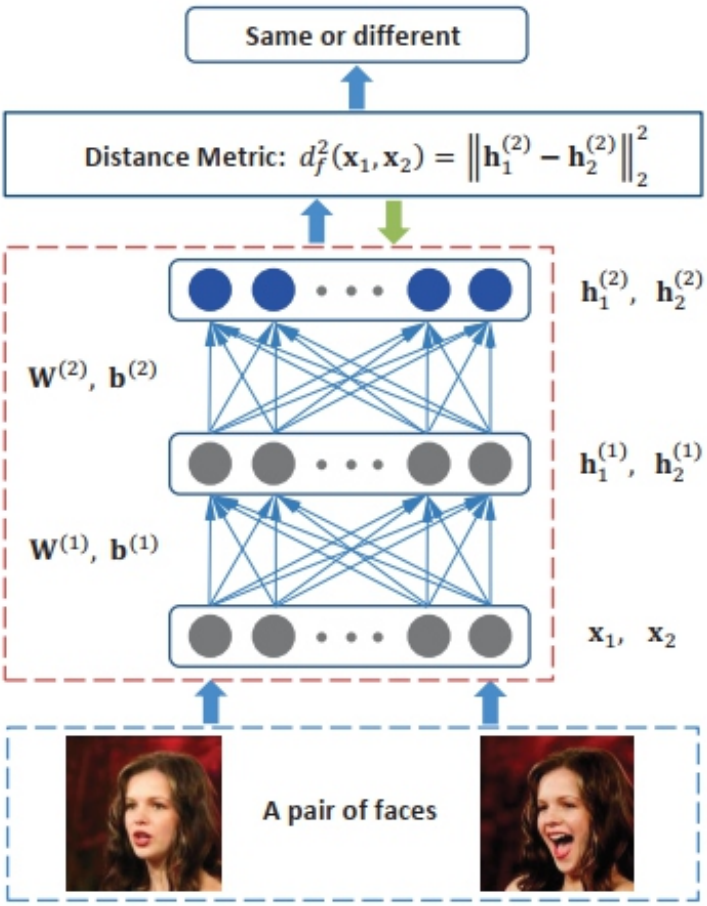


Figure 2.17: The flowchart of the proposed DDML method for face verification. Given a pair of face images x_1 and x_2, we map them into the same feature space $\mathbf{h}_1^{(2)}$ and $\mathbf{h}_2^{(2)}$ by learning a set of hierarchical nonlinear transformations, where the similarity between their outputs at the top level of the network is computed to determine whether the pair is from the same person or not. [101].

Hu et al. [61] discussed that the most existing metric learning methods performed to learn only one Mahalanobis distance metric from a monocular feature description for each facial image and cannot make utilize of various feature description directly. In several face-related topics, it can be easily extract various features for a facial image to perform more complementary data information, and it is eligible to learn distance metrics from these various features so that more effective information data can be extracted than those extracted from single features. To get this, they presented a large-margin multi-metric learning (LM^3L) approach for facial and kinship verification, which together learns various global distance metrics by which the correlations of various feature descriptions of each sample are enlarged, and the distance of each positive pair is minimal than a low threshold and that of each negative pair is bigger than a high threshold. To better extract the local data structures of facial images, they also proposed a local metric learning (LML) and a local large-margin multi-metric learning (L^2M^3L) approaches to learn a group of local metrics for face and kinship verification.

Qin et al. [131] proposed to learn and predict with gender-unknown kin relations. To address this issue, they presented a novel heterogeneous similarity learning (HSL) approach.

Encouraging by the fact that various kinship relations may not only share several common genetic traits but also have its own inherited characteristics from parents to offspring. They goal to learn a similarity metric by which the commonality among various kinship relations were kept and the geometry of each relation was saved, simultaneously. They further extended a multi-view HSL approach by optimal merge of the similarity methods from various feature descriptions, such that the integrally knowledge in multi-view kinship data information can be leveraged to obtain a better refined information. Fig. 2.18 shows illustration of pipeline of (a) the gender-fixed kinship verification and (b) the proposed heterogeneous kinship verification.

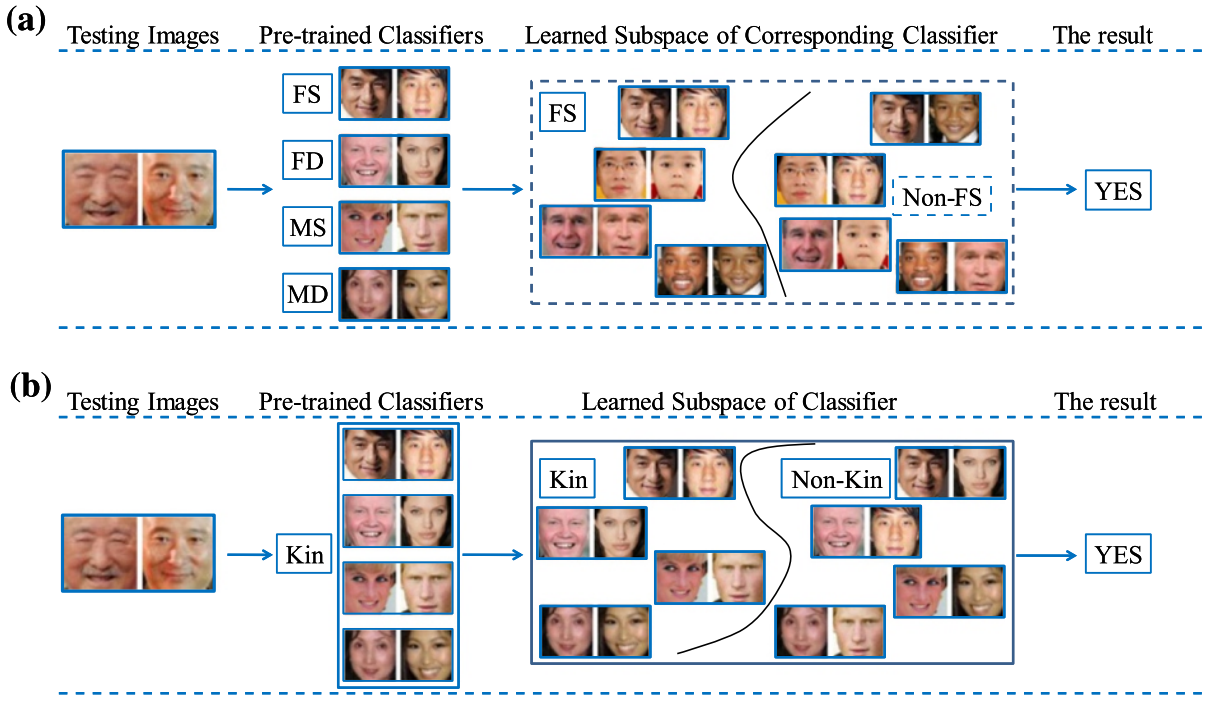


Figure 2.18: Pipeline of (a) the gender-fixed kinship verification and (b) the proposed heterogeneous kinship verification. For (a), one first train a classifier for each kind of kin relationship. When a pair of testing samples come in, one first pick out the corresponding classifier according to the genders of the query samples. Then, the selected classifier is conducted to find the real label of the testing samples; For (b), one first train a classifier for heterogeneous kin data. When a pair of testing samples come in, a fine classification is conducted to find the real label of the testing samples by using the trained classifier [131].

Zhao et al. [193] discussed that the related work methods focused either on discovering hand-crafted feature (shallow feature) descriptions to represent the facial or on learning the Mahalanobis metric distance to compute the similarity between face images. Instead, they proposed a new Multiple Kernel Similarity Metric (MKSM), by which, unlike from the Mahalanobis metric, the similarity calculation is essentially based on an inherent nonlinear feature transformation space. The general MKSM is a weighted collection of basic similarities and therefore have the capacity for features fusion. The essential similarities are derived from base kernels and local texture features, and the weights are resulted by solving a mannered linear programming (LP) problem that derived from a Large margin (LM) criterion. Furthermore, the LM criterion not only saves the generalization

on unseen data samples when the training data is very small, but also leads to sparsity in the weights vector in which turn boost the effectiveness at the prediction phase.

2.7 Convolutional deep learning-based kinship verification

In this section, we presented the state of the art methods were based essentially on deep convolutional neural networks to the problem of kinship verification.

Zhang et al. [188] proposed to learn high-level features descriptions for kinship verification based on deep convolutional neural networks. Their method is end-to-end, with non complex pre-processing often utilized in traditional approaches. The high-level features descriptions are resulted from the neuron activations of the last hidden layer, and then fed into a softmax classifier to make verifying the kinship of two facial images. Considering the significance of face key-points, they also extracted key-points-based features descriptions for kinship verification. Fig. 2.19 shows CNN-basic architecture.

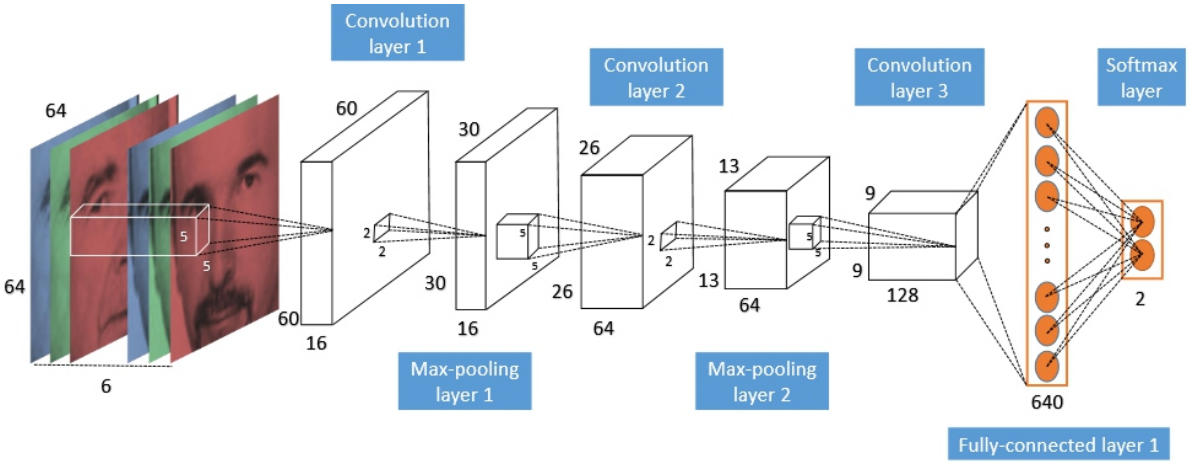


Figure 2.19: The architecture of basic CNN for kinship verification. For all layers,the length of each cuboid is the map number, and the width and height of each cuboid are the dimension of each map. The inside small cuboids and squares denote the 3D convolution kernel sizes and the 2D pooling region sizes of convolutional and pooling layers. The input is a pair of RGB images and the output is a two-value label [188].

Kohli et al. [79] made a human study that performed to understand the abilities of human perception and to identify the discriminated regions of a facial that smooth kinship cues. The visual stimuli offered to the participants define their capability to identify kinship relationship utilizing the total facial as well as particular facial parts. The influence of participant age and gender and kinship pairs of the stimulus is studied utilizing quantitative measurement such as accuracy, discriminability index d', and perceptual information entropy. Using the information data acquired from the human study, a hierarchical kinship verification via representation learning (KVRL) approach was used to better learn the description of various facial parts in an unsupervised manner. They proposed a new method for feature description named as filtered contractive deep belief networks (fcDBN). The proposed feature description combnes relational information data

present in facial images using filters and contractive regularization penalty. A compact description of face images of kinship was used as an output from the learned model and a multi-layer neural network was used to verify the kinship accurately. A novel WVU kinship database was collected, which comprises of several facial images per person to make easier kinship verification. The results conclude that the proposed deep learning approach (KVRL-fcDBN) outperforms the state-of-the-art kinship verification accuracy on the WVU kinship dataset and on four existing benchmark databases. Moreover, kinship information data was utilized as a soft biometric quality to increase the performance of facial verification by product of likelihood ratio and support vector machine based methods. Fig. 2.20 shows the KVRL-fcDBN approach for kinship verification.

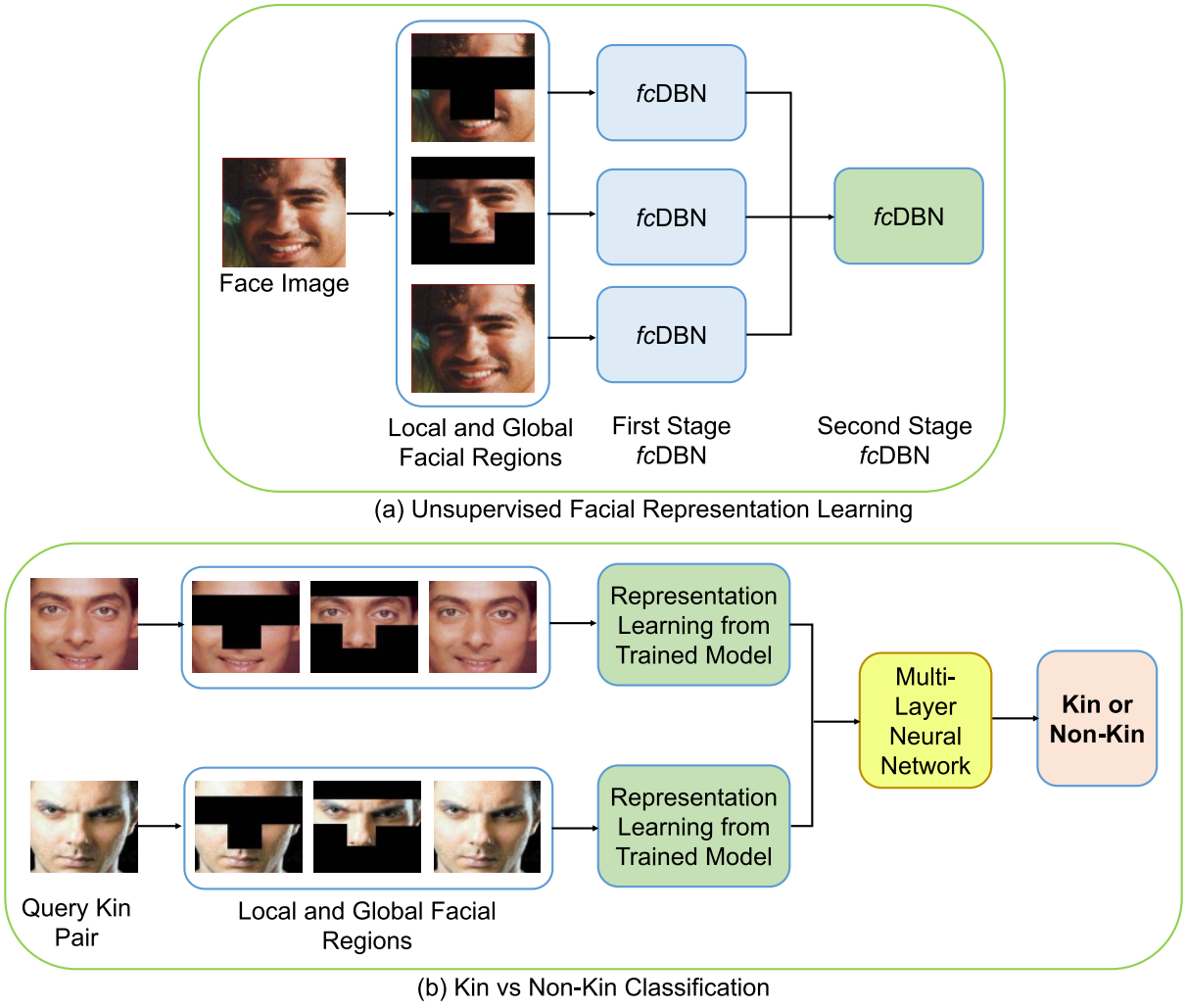


Figure 2.20: hierarchical kinship verification via representation learning (KVRL-fcDBN) framework. In the first stage of 2.20(a), representations of individual regions are learned. A combined representation is learned in the second stage of Fig. 2.20(a). Fig. 2.20(b) shows the steps involved in kin vs non-kin classification. [79].

Kohli et al. [80] proposed a novel deep learning approach for facial kinship verification in unconstrained videos utilizing a new Supervised Mixed Norm AutoEncoder (SMNAE). This novel autoencoder formularization presents class-specific sparsity includes in the matrix weight. The proposed three-stage SMNAE based facial kinship verification approach uses

the learned spatio-temporal description in the still video frames to verify the kinship in a pairs of videos. A novel kinship video (KIVI) dataset of more than 500 persons with variations due to occlusion, pose, illumination, ethnicity, and expression was gathered for their research. It includes basically of 355 positive kin video pairs with over 250 000 image frames. The efficiency of the proposed approach was applied and performed on the KIVI dataset and six other existed facial kinship datasets. On the KIVI dataset, SMNAE obtained video-based kinship verification accuracy of 83.18% which is at least 3.2% better than the existed methods. The approach was also tested on six publicly available facial kinship databases and achieves the best reported results. Fig. 2.21 illustrates the proposed kinship verification framework.

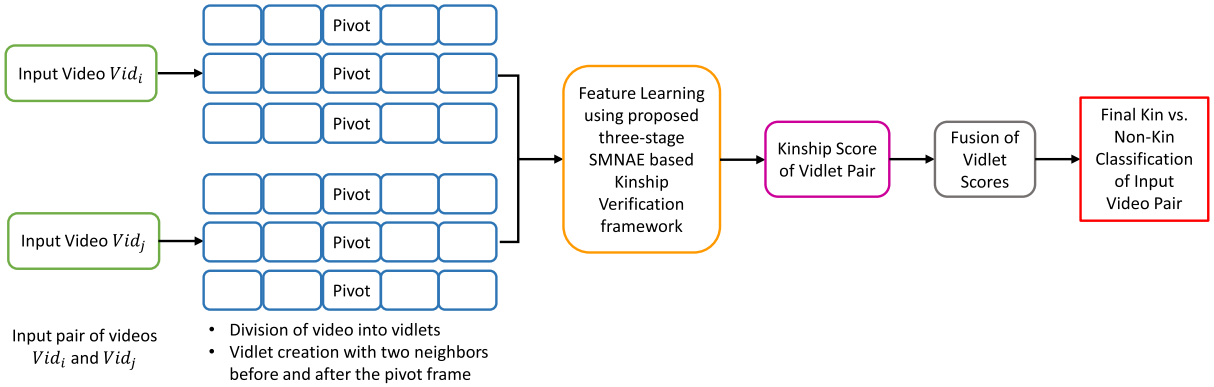


Figure 2.21: kinship verification framework: two input videos are divided into non-overlapping vidlets and for every vidlet, SMNAE features are extracted. The final video classification is performed by fusion of all the vidlet pair scores [80].

Yan et al. [178] presented an approach for face kinship verification, which utilizes an attention network to concentrate on obtaining discriminative information of local facial regions. Unlike most existed methods that use low-level features descriptions for kinship verification, they introduced an attention mechanism in the deep network to obtain high-level traits for face description. They also proposed a self-supervised method to orientate the attention network. Furthermore, they at random include a mask to five face features regions of each facial to get more help the network to focus on obtaining more effective data information at these group of regions. The visualization of the overall process illustrated in Fig. 2.22.

Wang et al. [155] proposed a towards-young cross-generation framework for efficient facial kinship verification by combining the two factors age and identity divergences. Fig. 2.23 shows Illustration of the cross-generation generative kinship verification framework. Furthermore, they explored a conditional generative method to force in an intermediate space to connect each pair. Moreover, they could obtain more discriminative features over deep networks with a newly-designed Sparse Discriminative Metric Loss (SDM-Loss), which was exploited to include the positive and negative data information.

Zhang et al. [189] show that the most existing approaches for facial kinship verification could be subdivided as handcrafted features-based shallow learning methods and

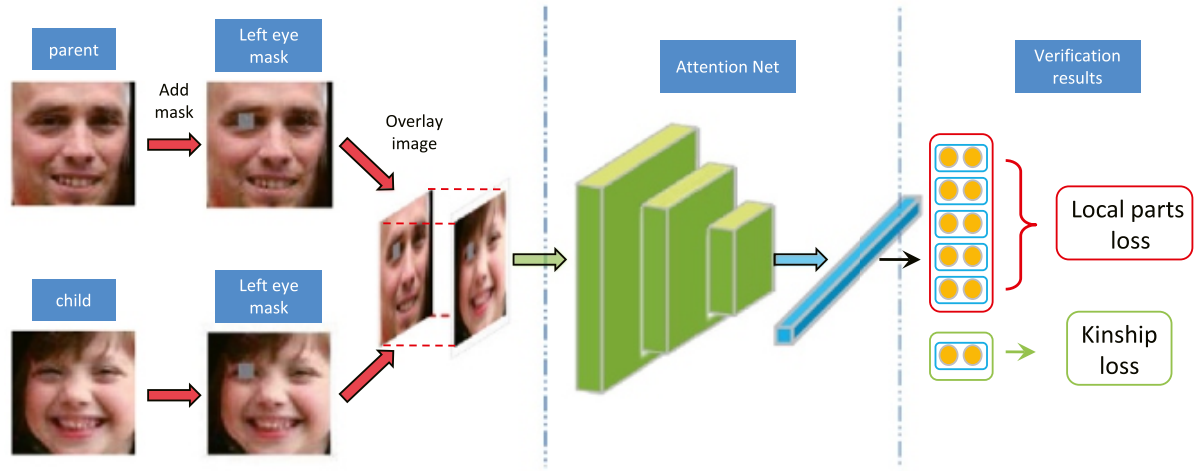


Figure 2.22: Flow chart of the experimental process. First cover a mask of the same local part for one image pair, then superimpose the two images together as an input to the network. The network outputs two different labels, where the local parts label records the location of the added mask, and the kinship label reports the verification result [178].

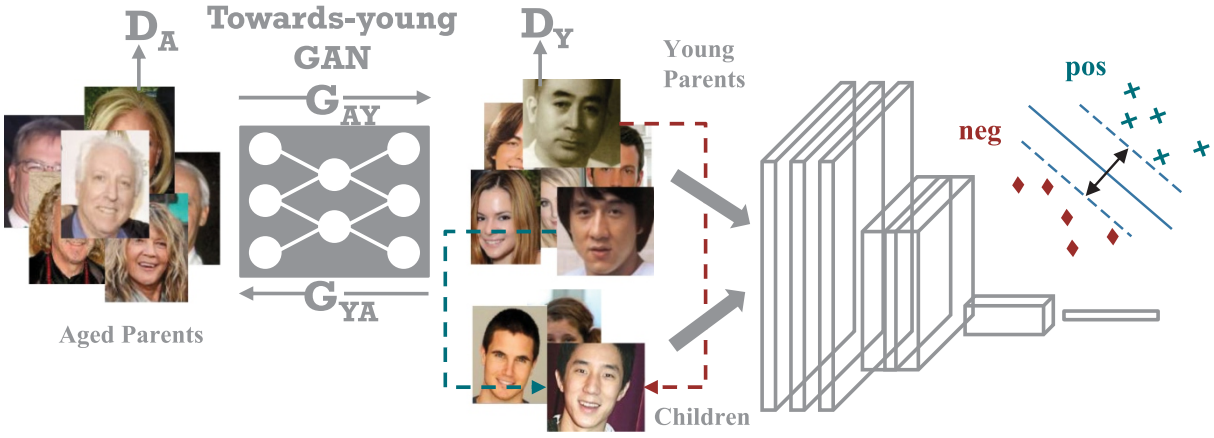


Figure 2.23: Illustration of the cross-generation generative kinship verification framework. The towards-young generative model is proposed to first generate young parents from its input old images and then the second stage network deal with the identity variation for a family pair with age gap mitigated and a newly-designed Sparse Discriminative Metric Loss (SDM-Loss), which is exploited to involve the positive and negative information [155].

convolutional neural network (CNN)-based deep-learning methods. Furthermore, these approaches are still deal the challenging problem of identifying kinship traits from face images. Moreover, the reason is that all the family ID identifiers information data and the distribution diverse of pairwise kinship from facials are seldom taken into consideration in facial kinship verification problems. Therefore, a family ID-based data adversarial convolutional network (AdvKin) approach mainly focused on effective Kinship traits was proposed for both kinds of kinship databases (large-scale and small-scale facial kinship verification). The advantages of their proposed framework are four-fold: 1) for kinship relationships detection, a simpler yet efficiency self-adversarial paradigm based on a negative maximum mean discrepancy (NMMD) loss was proposed as attacks in the first (FC-1) fully connected layer; 2) a pair-wise contrastive loss and family ID-based softmax loss are then together formulated in the second and third (FC-2 and FC3) fully connected layer, respectively, for supervised strategy training; 3) a two-stream network scheme with residual connections

was proposed in AdvKin; and 4) for more fine-grained deep kinship traits augmentation, an ensemble of patch-wise AdvKin networks was proposed (E-AdvKin). Fig. 2.24 illustrated the two-stream shared AdvKin approach.

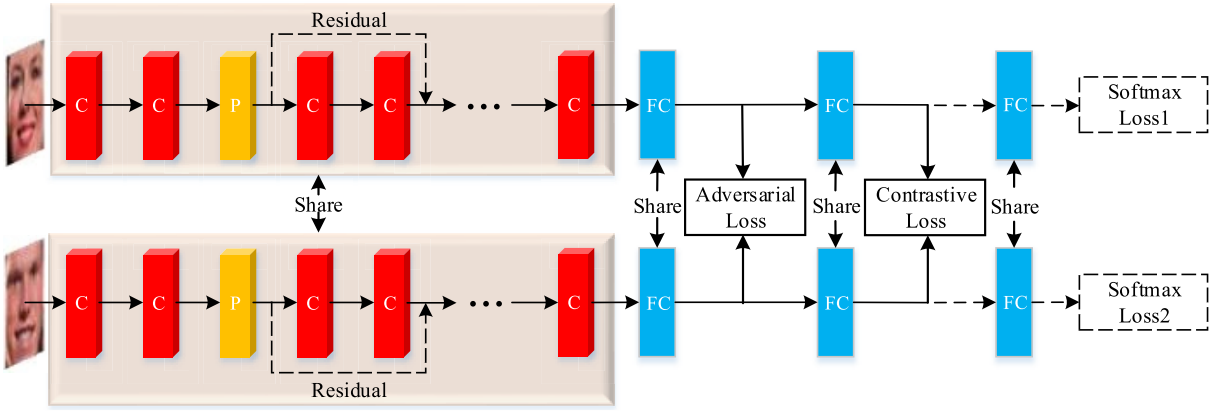


Figure 2.24: Pipeline of the two-stream shared AdvKin approach. C denotes convolution layer, P denotes pooling layer, and FC denotes fully connected layer. Note that the parts (i.e., residual connection versus SL layer) indicated by dashed lines are specifically added for large-scale kinship verification tasks [189].

Wang et al. [157] proposed a effective sampling approach to obtain the most efficient negative samples via deep reinforcement learning for facial kinship verification. Fig. 2.25 shows the deep convolutional architecture utilized for kinship verification. Unlike the most previous facial kinship verification approaches which focus basically on extracting discriminative features with the random sampling paradigm, they developed a deep reinforcement learning approach to obtain samples which are more suitable for learning discriminative kinship traits, so that the total performance could be enhanced. Furthermore, their approach utilizes two sub-networks to tackle the facial kinship verification problem: one DQN-based sampling architecture to filter the negative samples, and one multi-layer convolutional architecture to verify the kinship and make the final decision.

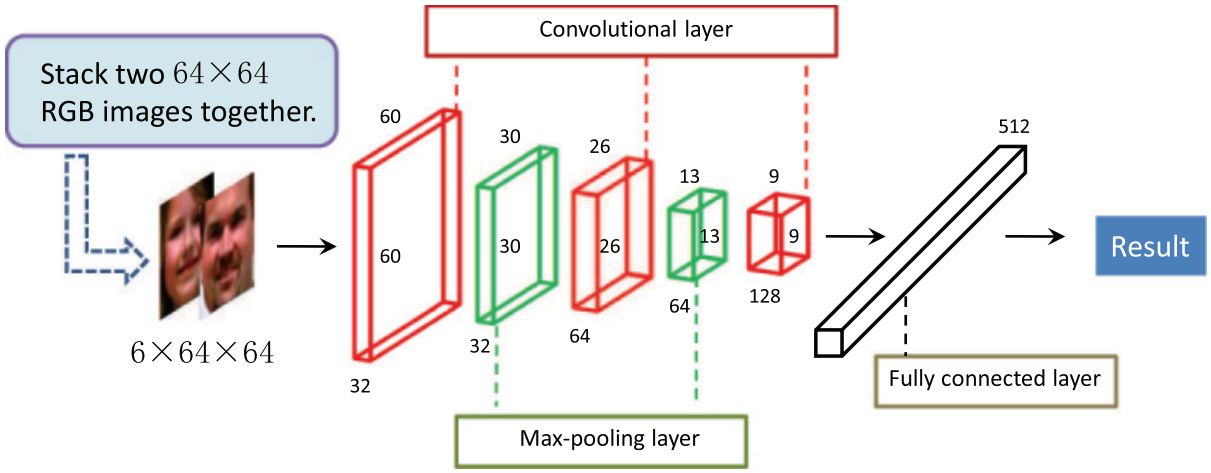


Figure 2.25: The figure details the structure of the deep convolutional network used for kinship verification. Each layer included in the network is shown in the figure. The dimension marked in the picture is the size of the data after passing through each layer. [157].

Yan et al. [177] proposed a deep relational network in which extracts multiple scale

information data of face images for kinship verification. Unlike the most previous deep learning based face kinship verification approaches which mostly used convolutional neural networks to obtain holistic features, they presented a deep model to obtain face kinship relationship from local parts. For each input pair of facial images, their method utilized two convolutional neural networks which share parameters to extract various scales of features, by which expected to give global contextual data information of facial images. They subdivide a set of traits at the one scale into multiple collections, where various collections capture information data of various local parts. For each face pair of local feature collections which are extracted from the same scale and position, they proposed a relation network to source their relationship, and utilized a verification model to infer the kinship relation based on the results of local relations from various facial parts. Fig. 2.26 shows the detailed structure of the deep relational network for facial kinship verification.

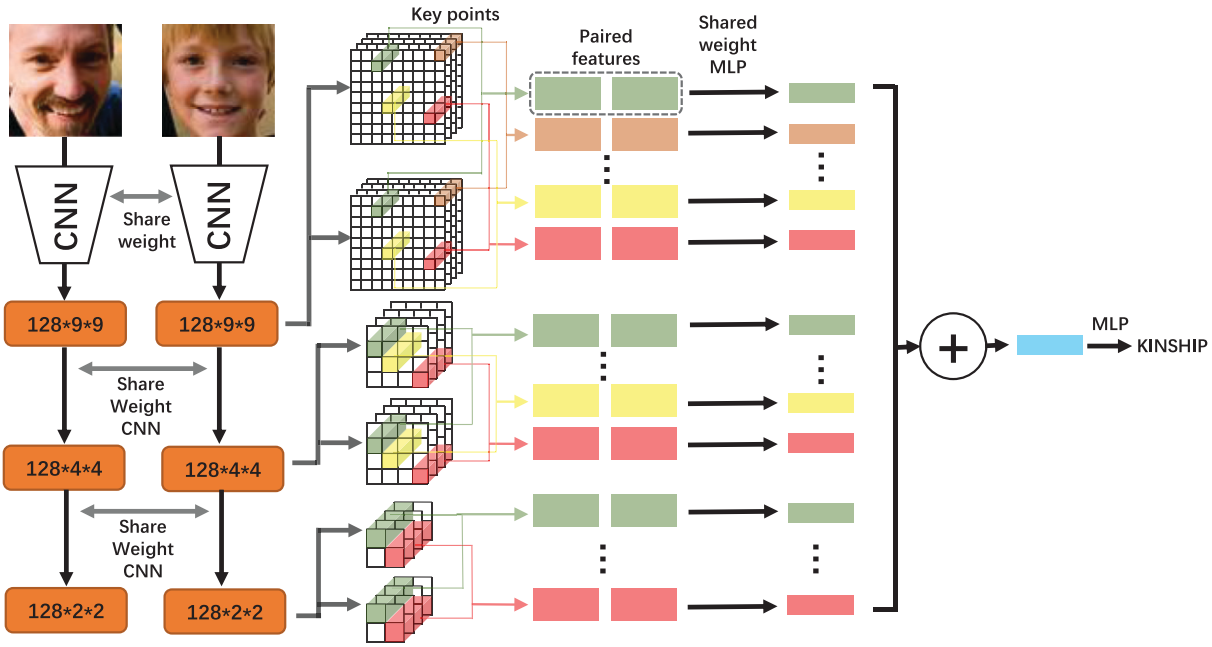


Figure 2.26: The detailed structure of the deep relational network. The network takes a pair of face images as the input. It first uses two convolutional neural networks with shared parameters to transform facial images into three scales of features. These features are only generated by partial face image regions due to different convolutional kernel sizes. These three scales of features provide both local and global information of face images. Features from two faces are concatenated together, and are processed by using a multi-layer perceptron with shared weights. The network then adds these features and uses another multi-layer perceptron to determine whether there is a kin relation or not for a given pair of face images [177].

We summarize in Table 2.1, the previous research works that show advance in the facial kinship verification field as well as their years of publication, results and benefiting of external training data for facial kinship verification or not.

Table 2.1: Review of facial kinship verification approaches. Outside Training column represents if an external face database was required for training the algorithm.

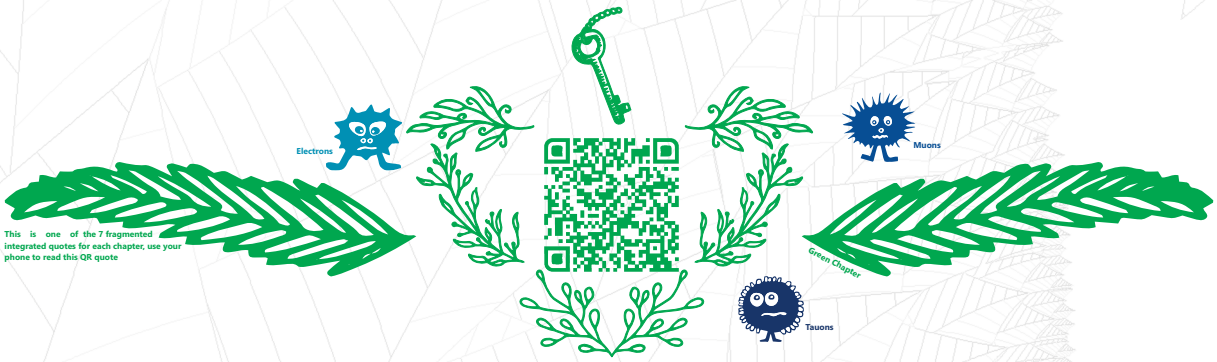
Year	Authors	Algorithm	Database	Accuracy (%)	Outside Training
2014	Lu et al. [103]	MNRML	KinFace-I	69.90	No
			KinFace-II	76.50	
	Lu et al. [174]	DMML	KinFace-I	72.25	
			KinFace-II	78.25	
			Cornell KinFace	73.75	
2015	Yan et al. [176]	MPDFL	KinFace-I	70.10	
			KinFace-II	77.00	
			Cornell KinFace	71.90	
			UB KinFace	67.30	
	Zhang et al. [188]	CNN-Points	KinFace-I	77.50	
			KinFace-II	88.4%	
2016	Kohli et al. [79]	KVRL+fcDBN	Cornell KinFace	89.50	Yes
			UB KinFace	91.80	
			KinFace-I	96.10	
			KinFace-II	96.20	
2017	Yan et al. [172]	NRCML	KinFace-I	66.30	No
			KinFace-II	78.70	
	Dibeklioglu et al. [37]	Method in [37]	KinFace-I	80.50	
			KinFace-II	82.30	
	Lu et al. [102]	DDMML	KinFace-I	83.50	
			KinFace-II	84.30	
			TSKinFace	84.15	
2018	Dawson et al. [32]	FSP	KinFace-I	76.80	
			KinFace-II	90.20	
			TSKinFace	88.60	
			Cornell KinFace	76.70	
	Liang et al. [92]	WGEML	KinFace-I	90.47	
			KinFace-II	82.80	
			TSKinFace	78.70	
	Kohli et al. [80]	Method in [80]	KinFaceW-I	96.90	Yes
			KinFaceW-II	97.10	
			Cornell KinFace	94.40	
			UB KinFace	95.30	
			TSKinFace	88.11	
2019	Moujahid et al. [114]	PML-COV-S	KinFace-I	88.20	No
			KinFace-II	88.20	
			UB KinFace	84.50	
	Dornaika et al. [39]	MNRML+SVM	KinFace-I	84.55	
			KinFace-II	86.90	
	Zhou et al. [195]	KML	KinFace-I	82.80	
			KinFace-II	85.70	
			Cornell KinFace	81.40	
			UB KinFace	75.50	
2020	Zhang et al. [189]	AdvKin	KinFace-I	78.70	
			KinFace-II	88.00	
			Cornell KinFace	75.00	
			UB KinFace	81.40	
	Wang et al. [157]	NESN-KVN	KinFace-I	78.60	
			KinFace-II	89.00	
	Yan et al. [177]	Method in [177]	KinFace-I	85.60	
			KinFace-II	88.80	

2.8 Conclusion

In this chapter, we introduced the general common challenges encountering facial kinship verification. Also, we explored the various measuring kinship characteristics which can be benefit from the facial kinship verification systems. Moreover, we described the general kinship verification framework and each essential components to make a decision. Furthermore, an overview was presented about facial kinship verification frameworks and we roughly categorise them into three types as it: Features learning-based kinship verification methods, Metric learning-based kinship verification methods and Convolutional deep learning-based kinship verification methods.

3 Kinship Verification from Face Images in Discriminative Subspaces of Color Components

Contents



3.1 Introduction

The majority of previous works on automatic verification of kinship have focused on the use of gray-scale information of face images, thus ignoring color information that can be an important clue for enhancing the verification performance. Biologically [144], the human traits' color can differentiate between individuals. For this reason, the chromaticity of the face parts, such as skin and eye color, is considered as significant imprint that distinguishes individuals. Furthermore, prior research indicates that color-texture provides useful information for face recognition compared with the luminance information. The experimental results in [147] show that the principal component analysis method using color information can improve the recognition rates compared with utilizing the luminance information only. The results in [183] show that color cues play a role in face recognition and their contribution becomes evident when shape cues are degraded. The results in [26] further elucidate that color cues can greatly ameliorate recognition performance compared with luminance-based features for dealing with low resolution face images. Other works from research community also demonstrate the efficiency of color-texture for face recognition [64, 97, 165, 181].

There exists a number of different color representations of images. Color spaces are the mathematical representation of color images in computer vision. The most frequently used image color space is RGB. However, RGB space has some limitations compared with other chrominance-luminance spaces, such as Luv and YCbCr spaces, since they are very close to human perception [18].

The purpose of this chapter is to investigate the efficiency of color-texture features extracted from facial images for the verification of kinship relations. Therefore, we consider the effect of seven color spaces, HSL, HSV, Lab, Luv, RGB, YCbCr and YUV. We compare the effectiveness of using different descriptors in gray-scale component versus their counterparts from color images in the seven spaces. More specifically, we use the color-texture characteristics from color images to encode the chrominance and the luminance information together. The contributions of this work are summarized as follows:

1. We propose an effective approach for kinship verification based on multiscale feature extraction projected into side information exponential discriminant analysis (SIEDA) subspace and fused various features using Logistic Regression (LR) scores fusion.

2. We investigate the efficiency of color-texture information through discriminative subspace using two-step learning approach, SIEDA and Logistic Regression, for automatic verification of kinship from facial images.

3. We evaluate various color spaces and descriptors on four benchmark kinship databases. Especially, each color channel of face image from a specified color space is projected

through the same implicit learned color channel subspace, and then all the channels information are fused to achieve better discrimination.

4. We study the complementarity of the different descriptors from the different color components.

The chapter is organized as follows: Our proposed method for color face kin verification is described in Section 3.2. The experimental data and setup are presented and results are discussed in Section 3.3. Finally, concluding remarks are given in Section 3.4.

3.2 Color-based face kinship verification

We aim to investigate the effectiveness of using color information by extracting face representations from luminance and chrominance components in several color spaces using different texture descriptors. Figure 3.1 shows our kinship verification scheme. The Parent-Child pair of face images is given as an input. First, the pair face images is cropped and normalized into an X × Y pixels. Then, we convert each pair of face images from (RGB) into different color spaces (e.g. HSV). Then, we extract the local features from each channel using multi-scale local descriptors of the considered color space (e.g. H, S and V). The encoded images are divided into K non-overlapping rectangular patches and each patch is summarized by a histogram. The histograms of different patches are concatenated to form a high dimensional feature vector. The features are then projected into ($SIEDA$) subspace. We compute the cosine similarity between the projected feature vectors of the pair for each color channel. We apply score fusion of the three scores resulting from the considered color space channels using Logistic Regression method (LR) [56]. Finally, the fusion score is compared to a threshold, set from the receiver operating characteristic (ROC) curve during performance evaluation, to decide whether the pair belongs to the persons from the same family or not. In the following we provide the details of the steps of our approach.

3.2.1 Color spaces

A color space is a mathematical representation of a set of colors. There are three popular color models: RGB (utilized in computer graphics), YUV or YCbCr (utilized in video systems) and CMYK (utilized in color printing). However, none of these color spaces is connected to the notions of hue, saturation and brightness. Therefore, other models, such as HSL and HSV, are invented to facilitate programming, processing, and end-user manipulation. All of the color spaces can be extracted from the RGB information provided by devices such as cameras and scanners.

In this work, in addition to RGB color space, we consider six other color spaces: HSL, HSV, Lab, Luv, YCbCr and YUV to examine the effectiveness of the color-texture

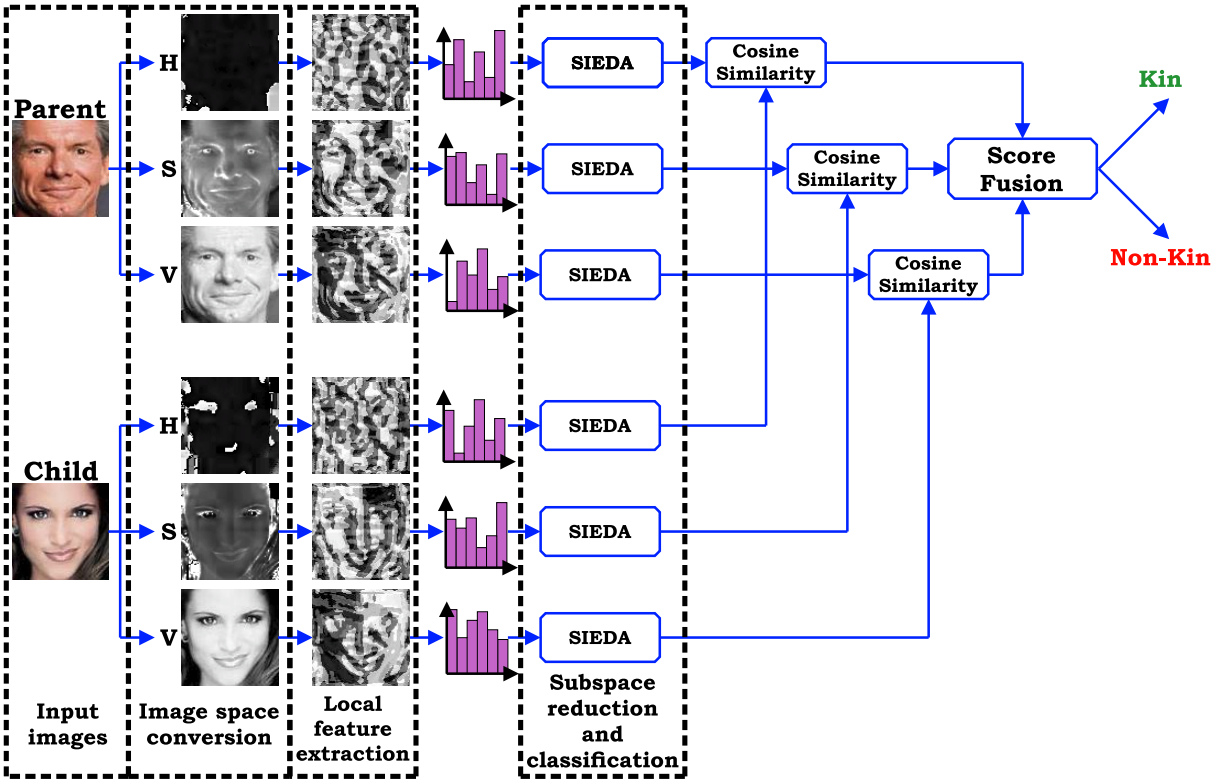


Figure 3.1: The proposed kinship verification approach.

information for kinship verification problem. The RGB color space is mostly used in computer graphics and displays. Red, green, and blue are the three primary individual components that can be mixed together to create a desired color. The colors are represented by a three-dimensional Cartesian coordinate system. On the other hand, HSV and HSL are the cylindrical-coordinate representation of every point in the RGB color space. These two color spaces avoid the RGB-like representation (the Cartesian cube) to achieve better perceptual description. Lab and Luv are two CIE (Commission Internationale de l'Éclairage) based color spaces. Both are founded on the CIE system of color measurement, which is based on human vision. The component L stands for lightness and (a, b and u, v) for the two color components green-red and blue-yellow. YCbCr and YUV are used for image pipeline specially in video and digital photography. The component Y in the both spaces represents the luminance, or intensity, which is appropriate for black and white display devices. Cb and Cr are the blue-difference and red-difference chrominance components, respectively in YCbCr space color. In YUV space color, U and V represent chrominance components.

3.2.2 Feature extraction

To describe face images, we extract three popular local texture descriptors: the Binarized Statistical Image Feature (BSIF) [77], the Local Phase Quantization (LPQ) [120] and the Co-occurrence of Adjacent Local Binary Patterns (CoALBP) [117]. To increase the

verification rate, we extract the three descriptors at multiple scales by varying the values of their parameters, i.e., W the filter size of BSIF; M the window size of LPQ; and S the number of directions and R the radius of circle of CoALBP. As an illustrative example, Figure 3.2 shows the application of LPQ and BSIF on the gray-scale, H, S and V color components.

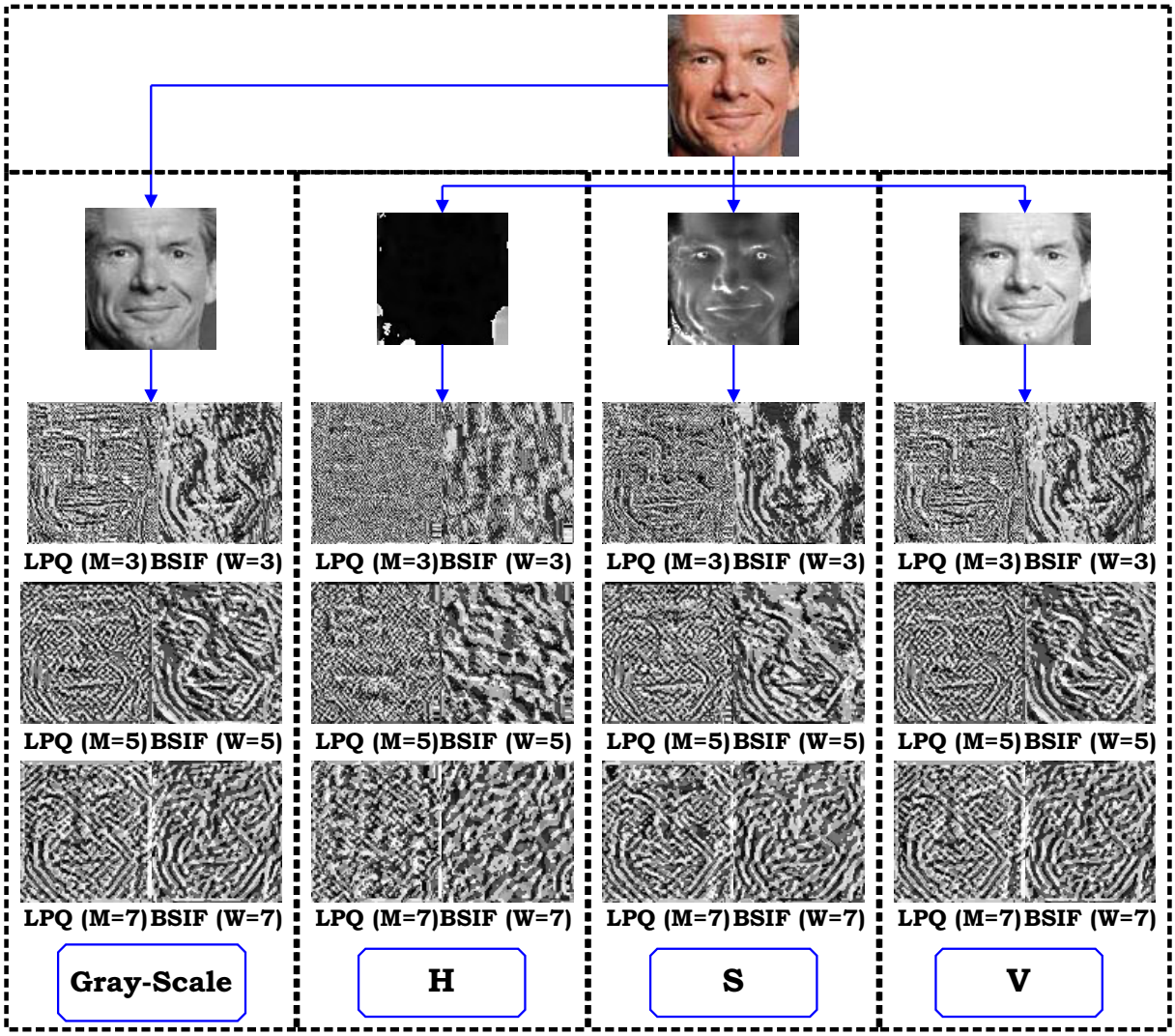


Figure 3.2: Illustrative example of application of Local Phase Quantization (LPQ) and Binarized Statistical Image Features (BSIF) on gray-scale and H, S and V color components of a face image.

3.2.3 Side-Information based Exponential Discriminant Analysis (SIEDA)

In order to enhance the discrimination of the extracted features, supervised methods, such as linear discriminant analysis, are usually applied to transform these features into (sub)space where their separability becomes easier. Supervised methods require the availability of class information for each sample. In the case of LDA [192] this information is used to compute the within class scatter matrix (M_w) and the between class scatter matrix (M_b). However, in kinship verification the only available information at training

stage is whether a pair of images belongs to the same family or not. Usually, the identity (or label) of each image is not available. In such case, refereed to as weakly labeled or weakly supervised, LDA is not applicable. To tackle this issue, Kan *et al.* [109] proposed a new approach where M_w and M_b matrices are computed using the weak labels (or side-information). In this case, the positive image pairs are utilized to calculate the within class scatter matrix and the negative image pairs are used to compute the between class scatter matrix. Let $P_{class} = \{(z_i, z_j) : k(z_i) = k(z_j)\}$ be the set of positive image pairs and $N_{class} = \{(z_a, z_b) : k(z_a) \neq k(z_b)\}$ the set of negative image pairs, where the image z is represented by the class label $k(z)$. The within-class and between-class scatter matrices of Side-Information based Linear Discriminant analysis (SILD) method are obtained as follows:

$$M_w^{sild} = \sum_{(z_i,z_j)\in P_{class}} (z_i - z_j)(z_i - z_j)^T \tag{3.1}$$

$$M_b^{sild} = \sum_{(z_a,z_b)\in N_{class}} (z_a - z_b)(z_a - z_b)^T \tag{3.2}$$

The optimization function for SILD is:

$$\begin{aligned} U_{opt}^{sild} &= argmax_U \frac{U^T M_b^{sild} U}{U^T M_w^{sild} U} \\ &= argmax_U \frac{\left|U^T (V_b^T \Lambda_b V_b) U\right|}{|U^T (V_w^T \Lambda_w V_w) U|} \end{aligned} \tag{3.3}$$

This problem is reduced to finding the eigenvectors V and the eigenvalues Λ of the matrix $M_w^{-1} M_b$.

$$M_w^{-1} M_b = V^T \Lambda V \tag{3.4}$$

One particular problem that is frequently confronted to when applying LDA-like methods is the small sample size problem of LDA-based facial recognition frameworks. The dimension of feature vectors is usually very high compared with the number of per class samples, which causes the singularity of the within-class scatter matrix of LDA and SILD. To overcome this limitation, Kan *et al.* [109] apply PCA [40] first to lower the dimension of feature vectors then SILD is applied. However, proceeding so has the inconvenient that the discriminative information included in the null space of M_w^{sild} are lost. To retain this useful discriminative information, Ouamane *et al.* [121] proposed the Side-Information Exponential Discriminant Analysis (SIEDA). Inspired by EDA [190], SIEDA replaces Λ_{wk}, the eigenvalues of M_w^{sild}, by $exp(\Lambda_{wk})$ and Λ_{bk} the eigenvalues of M_b^{sild} by $exp(\Lambda_{bk})$. Thus, the target function for SILD becomes:

$$\begin{aligned} U_{opt}^{sieda} &= argmax_U \frac{\left|U^T (V_b^T exp(\Lambda_b) V_b) U\right|}{|U^T (V_w^T exp(\Lambda_w) V_w) U|} \\ &= argmax_U \frac{U^T exp(M_b^{sild}) U}{U^T exp(M_w^{sild}) U} \end{aligned} \tag{3.5}$$

The projection matrix U_{opt}^{sieda} is composed by the most significant eigenvectors of $(exp(M_w^{sild}))^{-1}exp(M_b^{sild})$.

The advantages of SIEDA compared to SILD are threefold: firstly, from the properties of matrix exponential, the matrix $exp(M_w^{sild})$ is a full-rank matrix. Therefore, the null space of M_w^{sild} which includes the discriminant information in the solution of Equation (3.5) will be preserved. Secondly, a kernel method (diffusion) is utilized for the transformation of the nonlinear posed problem toward linear problem. Similarly to the kernel methods, the scatter matrices are transformed into a new space by the exponential function [124]. Finally, the target function for SILD is to maximize the distance of the between-class and minimize the distance of the within-class. The corresponding trace of scatter matrices, $trace(M_b^{sild}) = \Lambda_{b1}+\Lambda_{b2}+\cdots+\Lambda_{bn}$ and $trace(M_w^{sild}) = \Lambda_{w1}+\Lambda_{w2}+\cdots+\Lambda_{wn}$, ensures the satisfaction of the two conditions. Thus, the larger of the ratio $\Lambda_{bk}/\Lambda_{wk}$ is, the higher of the discrimination power. From the matrix exponential properties, we have: $trace(exp(M_b^{sild})) = exp(\Lambda_{b1}) + exp(\Lambda_{b2}) + \cdots + exp(\Lambda_{bn})$ and $trace(exp(M_w^{sild})) = exp(\Lambda_{w1}) + exp(\Lambda_{w2}) + \cdots + exp(\Lambda_{wn})$. Therefore, the ratio $exp(\Lambda_{bk})/exp(\Lambda_{wk})$ is bigger than $\Lambda_{bk}/\Lambda_{wk}$. Thus, one can conclude that there is a difference in the spread scale through the within and between-class distances which leads to a better separation.

3.3 Experiments

In this section, we perform a number of experiments to evaluate the proposed kinship verification system and investigate the strengths of color-texture representations. In addition to their fusion, different color spaces are studied individually. Firstly, we present the benchmark databases used in our experiments. Then, we investigate the advantage of color-texture for kinship verification. Finally, we provide and discuss the results and compare the best ones with those of the state of the art.

Figure 3.3 shows samples of positive pairs and negative pairs from the kinship databases.

3.3.1 Parameter Settings

The performance of the proposed approach is evaluated on the same experimental protocol as the literature works [103, 172, 174], where five-fold cross-validation for kin verification is performed while keeping the number of pairs images roughly equal in all folds. This protocol is followed to ensure that our results are directly comparable to the state of the art. The negative pairs for kinship are generated randomly such that each image appears only once in the training set. The number of positive pairs and negative pairs is the same for both training and test.

To mitigate the effect of face normalization and to be consistent with several previous works, including [3, 35, 96, 101, 103, 168, 172, 174, 175], all feature descriptions are extracted

Figure 3.3: (Left) Positive pairs (with kinship relation) and (right) negative pairs. From top to bottom the pairs are from Cornell, TSKinFace, KinFaceW-I and KinFaceW-II databases, respectively.

from face images that are aligned and cropped using the position of the eyes to 64×64 pixels.

Regarding feature extraction, we use eight filters with different sizes W = 3, 5, 7 in the Multi-Scale Binarized Statistical Image Features (MSBSIF). In the Multi-Scale Local Phase Quantization (MSLPQ), the window size is M = 3, 5, 7. The choice of these scales is to depend on the size of the face image. Every face image is subdivided into 64 blocks, each of size 8×8 pixels. We use histograms of 256 bins to aggregate the local features extracted from each block. By concatenating the histograms of all the 64 blocks, a vector of dimension $3 \times 64 \times 256$ is obtained. For Multi-Scale Co-occurrence of Adjacent Local Binary Patterns (MSCoALBP), the extracted features from each face image are the combination of three histograms where the size is 3×1024, using the LBP+ operator with radius R = 1, 2, 4 and the identical directions specified by the distances S = 2, 4, 8.

3.3.2 Face kinship similarities in color spaces

To analyze the effect of different color spaces on kinship verification, we first perform a preliminary experiment on the face similarities. As shown in Figure 3.4, similarities between face pairs in different color components are computed. The similarity between the pair of feature vectors is computed using the normalized cosine distance:

$$sim(U, V) = \frac{\vec{u}.\vec{v}}{\|\vec{u}\| \, . \, \|\vec{v}\|} \tag{3.6}$$

The using of cosine similarity distance after discriminant analysis methods has an adventage comes from its connection to the Bayes decision rule, as the optimal used method is the Bayes classifier for decreasing the classification error [95].

For this particular example in Figure 3.4, the similarity of the negative pair in gray-scale

is larger than that of the positive one. When considering the similarity between the same parts in the components of HSV color space, the similarity of the negative pair is lower than that of the positive pair with a significant margin. This means that, for this example, the kinship verification is clearly easier in HSV color space. For a general analysis, we plot, in Figure 3.5, the distributions of the similarities of positive pairs and negative pairs on TSKinFace database. The positive and negative pairs similarities in gray scale are dense and their overlap is considerable while the distributions in RGB and HSV color spaces are spread and the overlap is less significant. This first experiment suggests that color spaces are more discriminative through subspace transformation method (i.e. SIEDA) for kinship verification from face images. In the subsequent experiments this hypothesis will be tested.

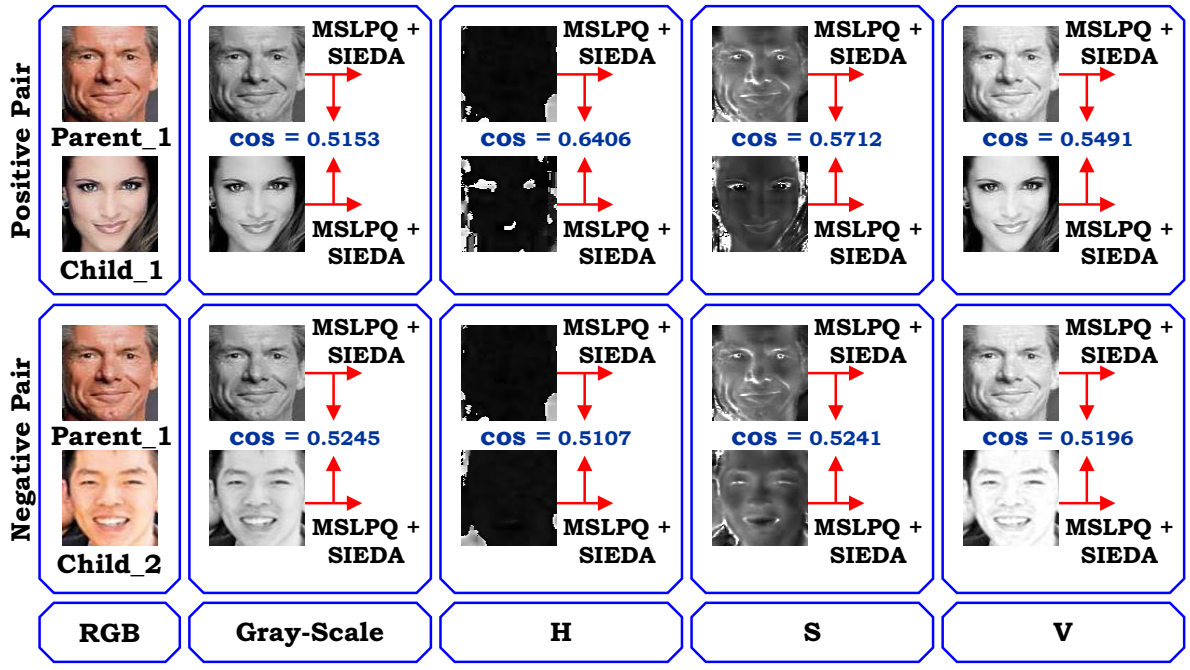


Figure 3.4: Example for computing the kin similarity between face image pairs in gray-scale and HSV components.

3.3.3 Results and discussion

To demonstrate the efficiency of SIEDA projection, we compare its performance to PCA and SILD methods as well as simple cosine scoring with no projection in kinship verification. We run the experiments on all the databases for gray-scale. Table 3.1 shows the obtained mean accuracy of kinship verification of different methods using MSLPQ features. These results show the superiority of SIEDA compared to other methods. This is mainly due to the enhanced discriminative power of SIEDA.

The experimental results of the three face descriptors, MSBSIF, MSLPQ and MSCoALBP separately as well as four fusion variants: Fusion1 (MSBSIF+ MSLPQ), Fusion2 (MSBSIF+ MSCoALBP), Fusion3 (MSLPQ+ MSCoALBP) and Fusion4 (MSBSIF+ MSLPQ+

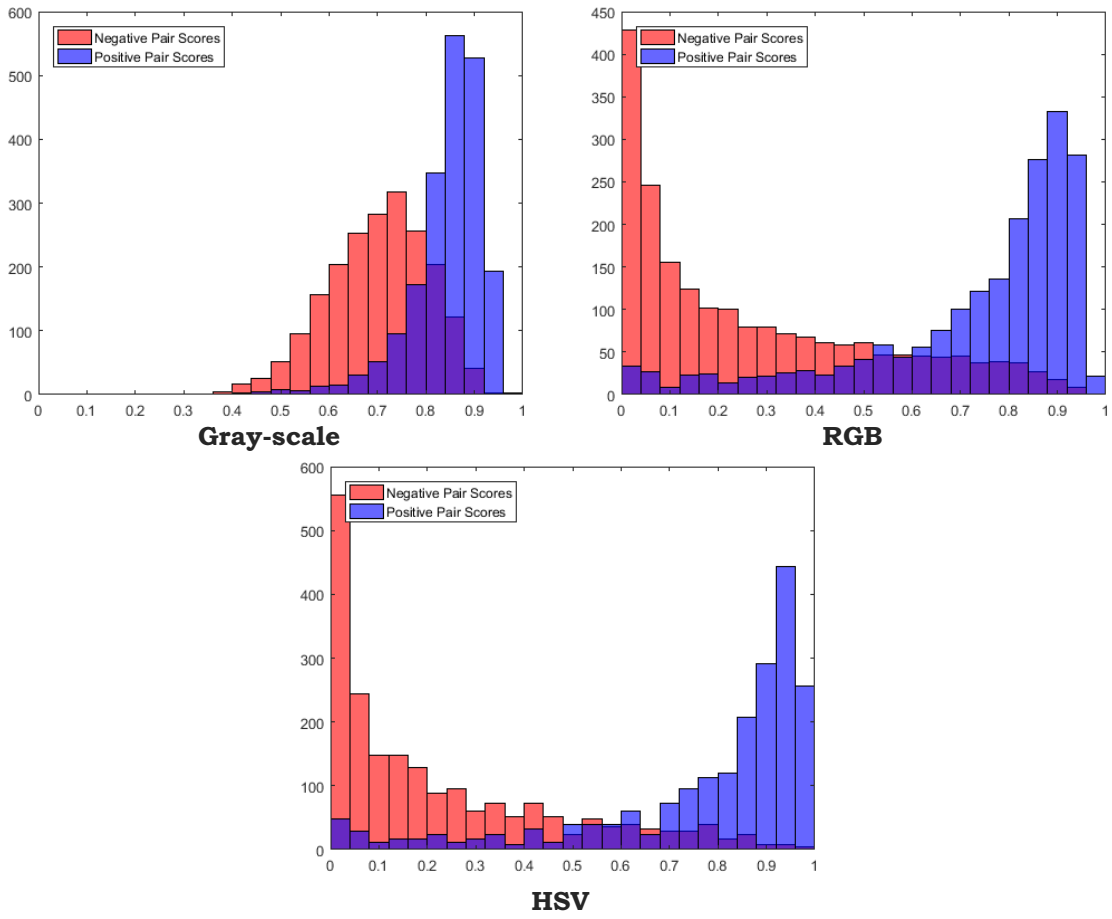


Figure 3.5: The distribution of similarities between the positive pairs (blue) and negative pairs (red) of TSKinFace Database using gray-scale images RGB and HSV color spaces.

Table 3.1: The mean accuracy (%) of kinship verification for simple scoring, PCA, SILD and SIEDA methods using MSLPQ for the gray-scale on each database.

Method / Database	Simple scoring	PCA	SILD	SIEDA
Cornell	68.13	69.93	73.11	76.94
TSKinFace	72.86	77.25	81.53	83.23
KinFaceW-I	71.75	73.27	74.24	76.30
KinFaceW-II	70.65	70.75	76.20	77.00

MSCoALBP) are depicted in Figures 3.6, 3.7, 3.8 and 3.9 for Cornell, TSKinFace, KinFaceW-I and KinFaceW-II, respectively. For detailed results of color spaces on the four databases, see Appendix D. The descriptor fusion is performed with the Logistic Regression method (LR) [56]. The performances of the different feature descriptors and their fusions are computed for the gray-scale as well as HSL, HSV, Lab, Luv, RGB, YCbCr and YUV color image representations.

Furthermore, these Figures clearly depict the fact that all the color spaces outperform the gray-scale in all configuration through the descriptors and databases. Table 3.2 shows the best accuracy from each color space in the experimental databases. The performance

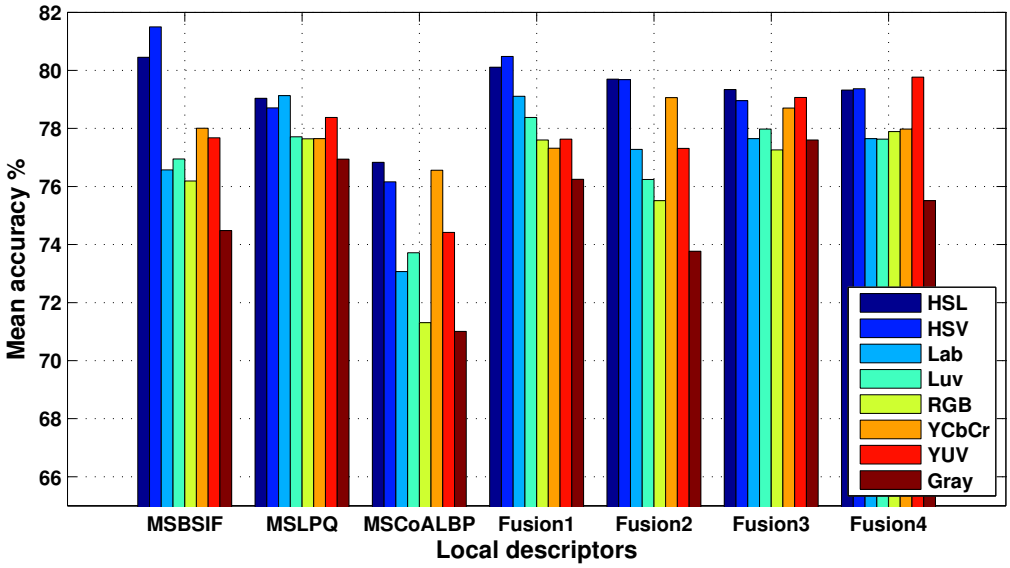


Figure 3.6: Mean kinship verification accuracy for individual descriptors and their fusion for the gray-scale and the seven color spaces on Cornell database.

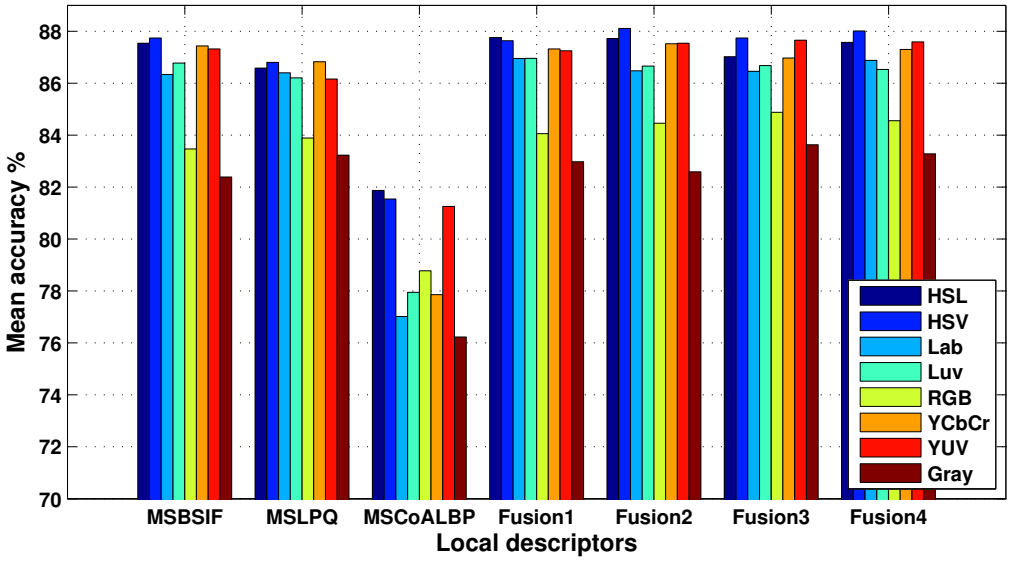


Figure 3.7: Mean kinship verification accuracy for individual descriptors and their fusion for the gray-scale and the seven color spaces on TSKinFace database.

is improved with variation between 2.4% and 9.6%. Moreover, the RGB, which is the most used color space for kinship verification, face recognition and analysis as well as many image processing applications, gives lower performance than the other spaces in most cases. The gain in performance using other color spaces than RGB is clearer and significant on Cornell, TSKinFace, KinFaceW-I and KinFaceW-II databases.

Generally on Cornell, KinFaceW and TSKinFace databases, MSBSIF performs the best compared to MSLPQ and MSCoALBP. On an other hand, each of Fusion1, Fusion2, Fusion3 and Fusion4 further improves the performances of individual descriptors through the color spaces and databases.

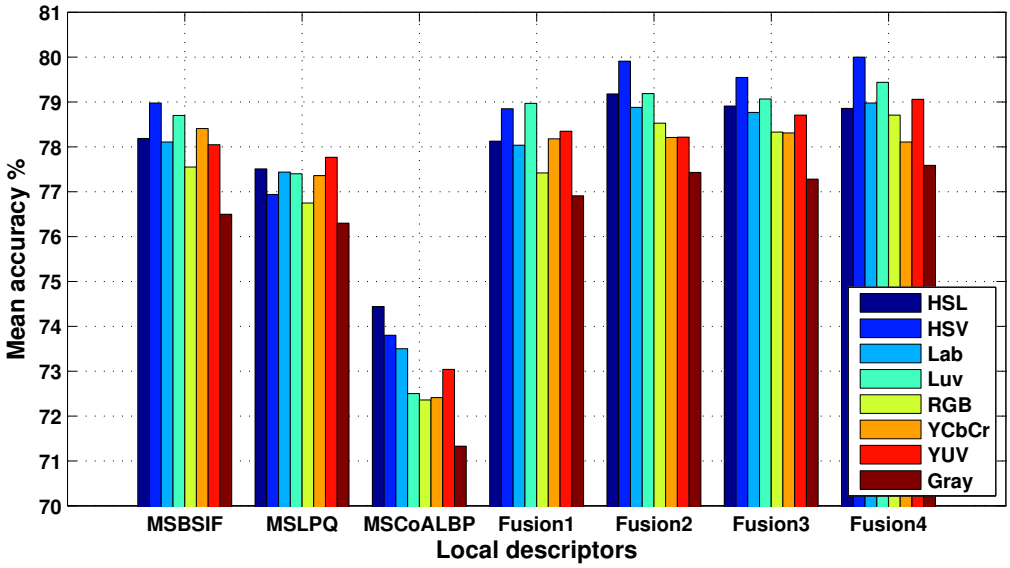


Figure 3.8: Mean kinship verification accuracy for individual descriptors and their fusion for the gray-scale and the seven color spaces on KinFaceW-I database.

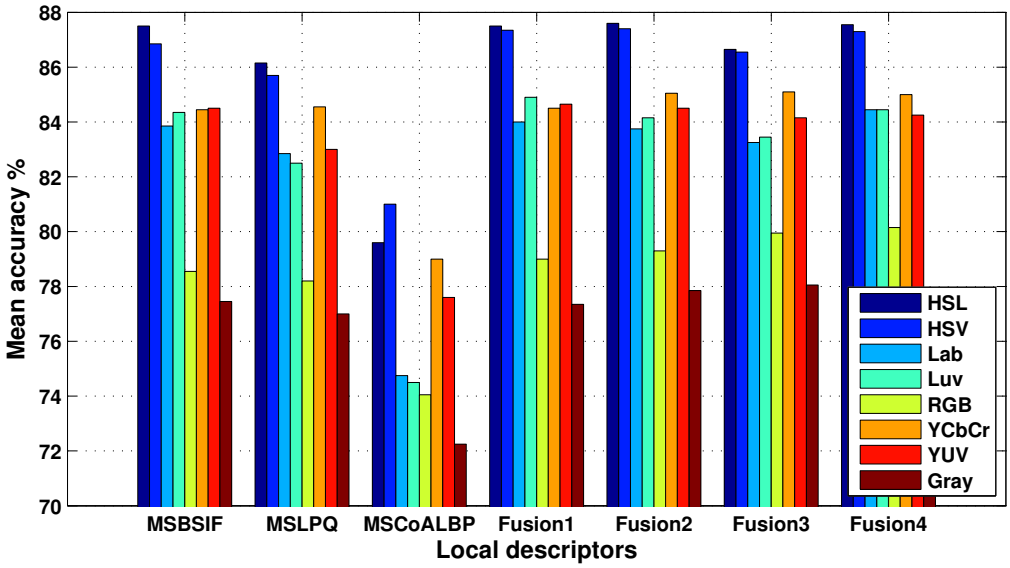


Figure 3.9: Mean kinship verification accuracy for individual descriptors and their fusion for the gray-scale and the seven color spaces on KinFaceW-II database.

Table 3.2: Best obtained accuracies (%) on each database for different color spaces and gray-scale.

Color space / Database	HSL	HSV	Lab	Luv	RGB	YCbCr	YUV	Gray
Cornell	80.45	81.50	79.13	78.38	77.89	79.06	79.77	77.60
TSKinFace	87.76	88.11	86.95	86.96	84.88	87.52	87.66	83.63
KinFaceW-I	79.18	80.00	78.98	79.44	78.71	78.41	79.06	77.59
KinFaceW-II	87.60	87.40	84.45	84.90	80.15	85.10	84.50	78.05

To check the performance of different kinship relations, we plot in Figures 3.10 and 3.11 the ROC curves for each relation on TSKinFace and KinFaceW-II databases. For clarity

purpose, only curves of Fusion4 for gray-scale and color spaces are depicted. The ROC curves confirm the previous remarks for individual relations. Color spaces are better than gray-scale and RGB performances are less than the other color spaces. Furthermore, HSV color space outperforms all color spaces in the most cases.

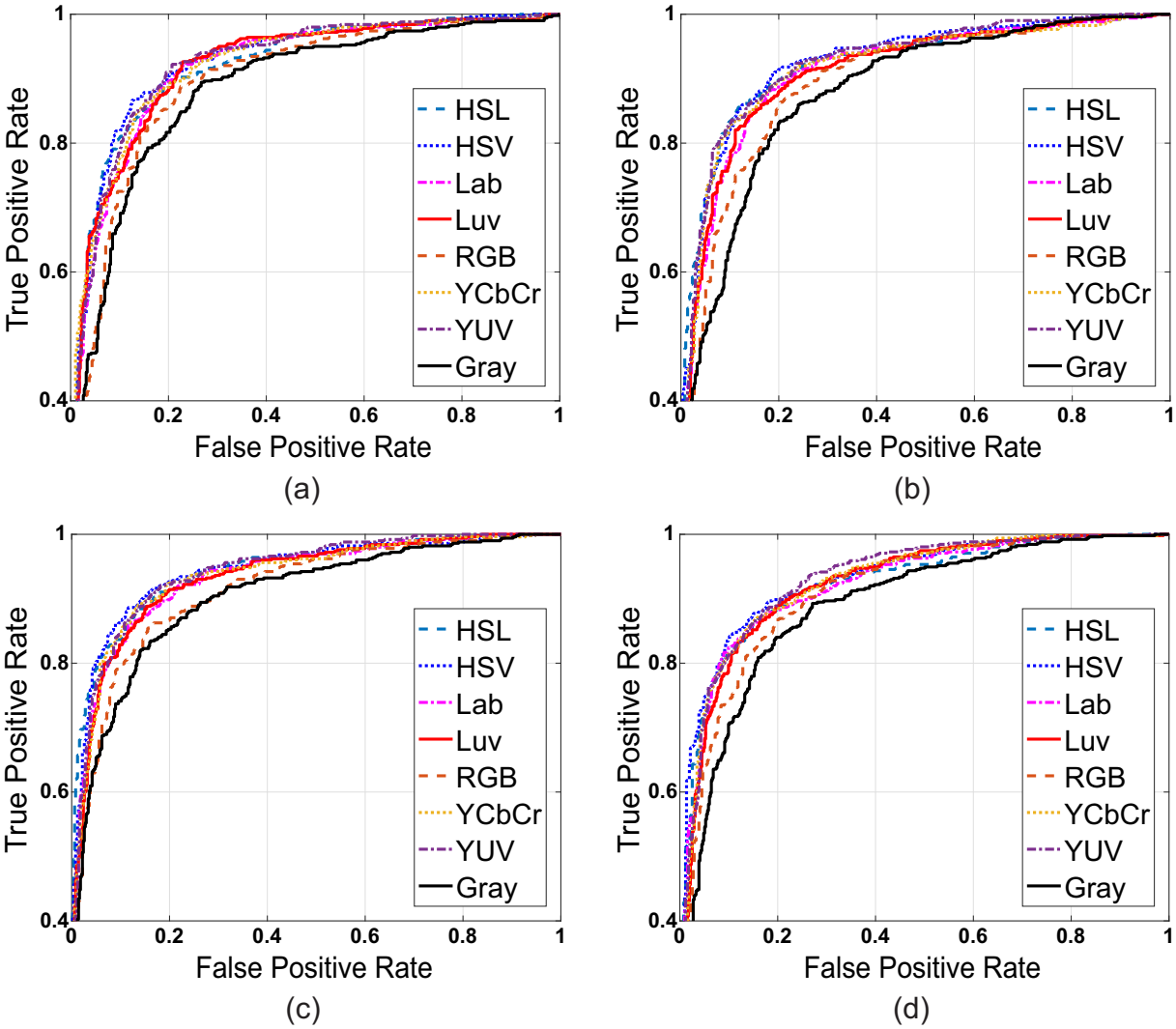


Figure 3.10: ROC curves of Fusion4 for (a) F-D, (b) F-S, (c) M-D and (d) M-S on TSKinFace database considering different color spaces.

3.3.4 Comparison with the results of the state of the art

The best kinship verification performances of our approach are achieved using: MSBSIF on Cornell, the three descriptors (Fusion4) on KinFaceW-I, and the two descriptors (Fusion2) on TSKinFace and KinFaceW-II databases. For color, verification rates of 81.50%, 88.11% and 87.60% are reported on Cornell, TSKinFace and KinFaceW-II databases, respectively. These results are compared with the state of the art in Table 3.3. From KinFaceW-I, we see that our proposed method achieve competitive performance with the cited kinship verification methods. The comparison reveals that our proposed color-texture

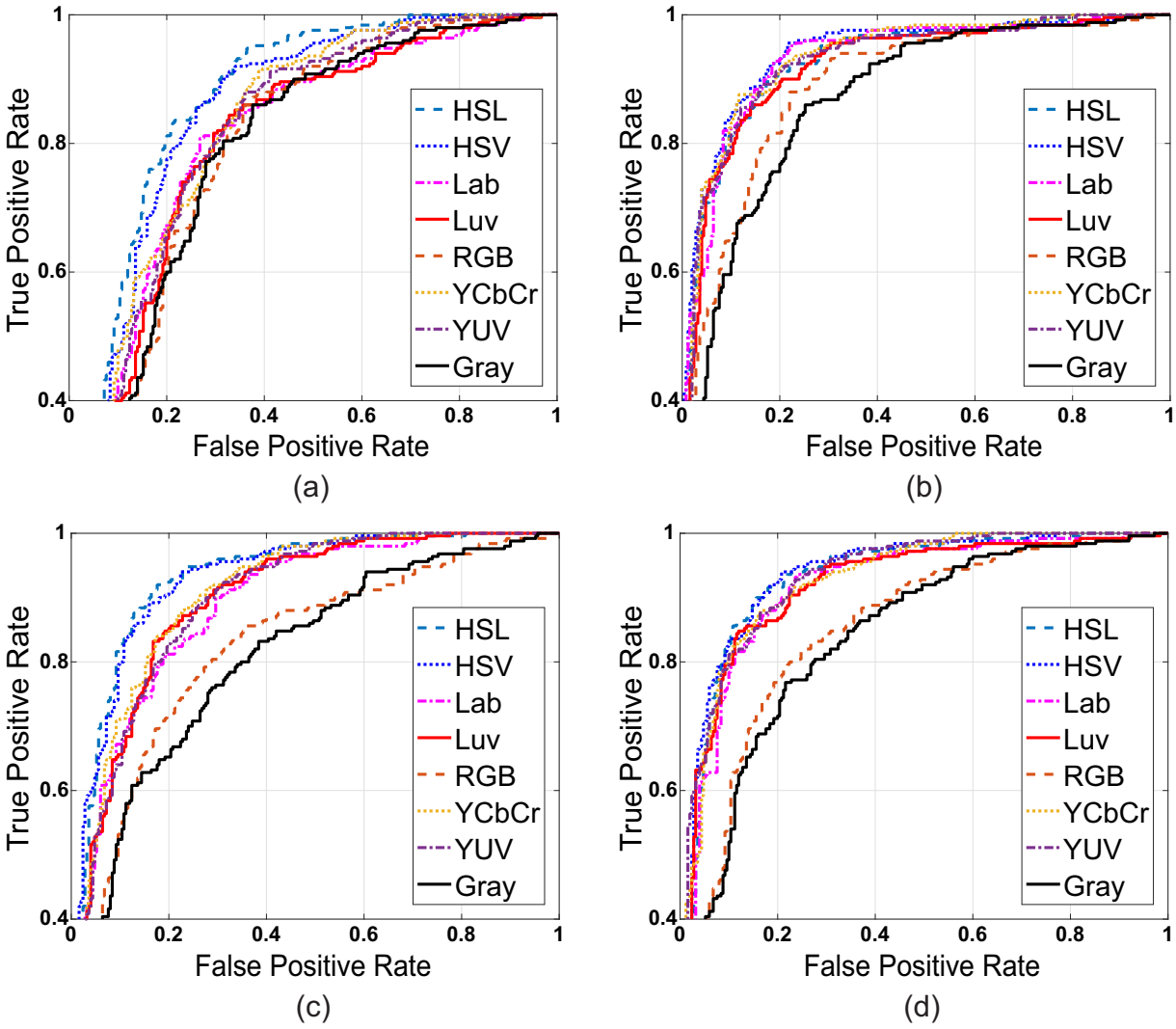


Figure 3.11: ROC curves of Fusion4 for (a) F-D, (b) F-S, (c) M-D and (d) M-S on KinFaceW-II database considering different color spaces.

analysis based method outperforms the recent state of the art on the three databases, Cornell, TSKinFace and KinFaceW-II. This demonstrates the effectiveness of using the face color-texture information in SIEDA subspace for kinship verification.

Our approach vs. Simple scoring approach [168] From the evaluation on TSKinFace database using the Gray and HSV spaces, we can see that our approach improves the performance about 9.6% and 6.9% using the Gray and HSV spaces, respectively. The work of Wu et al. [168] was based on simple scoring method, where only cosine similarity was used to compare kinship face images and obtained 74.03% and 81.19 % using Gray and HSV spaces, respectively. However, in our work, we used the two-step learning method (i.e. SIEDA/LR) to learn the kinship face images achieving better performances, 83.63% and 88.11% using Gray and HSV spaces, respectively. Furthermore, our approach used the three channels from each color space to learn three subspaces specific to each color space using SIEDA method. Then, the projected face pairs by three SIEDA subspaces are fused by Logistic Regression method. This is different from the work of [168] which

simply extracts features from separate channels without tacking redundancy, correlation and complementarily into account. Therefore, our approach considers and handles better these aspects.

Table 3.3: Comparison with state of the art.

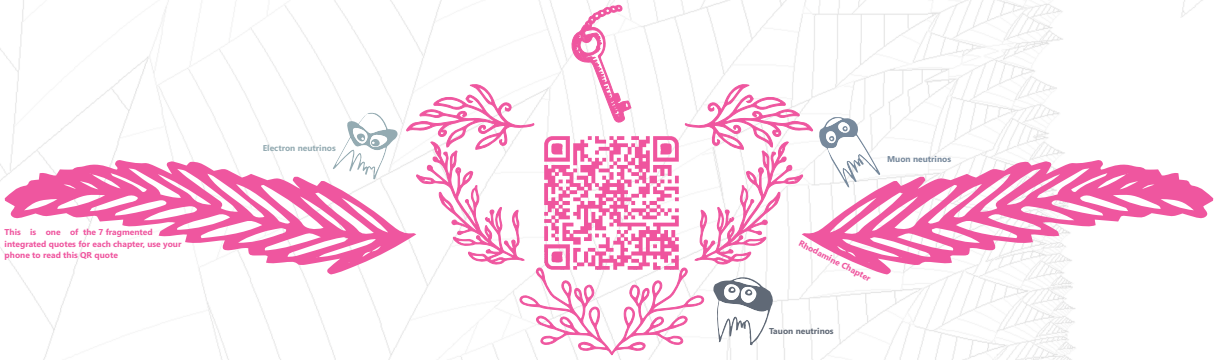
Author	Cornell	TSKinFace	KinFaceW-I	KinFaceW-II
Fang *et al.* [44] (2010)	70.67	/	/	/
Lu *et al.* [103] (2014)	71.60	/	69.90	76.50
Yan *et al.* [174] (2014)	73.50	/	72.00	78.00
Dehghan *et al.* [35] (2014)	/	/	74.50	82.20
Yan *et al.* [175] (2015)	71.90	/	70.10	77.00
Liu *et al.* [96] (2015)	/	/	73.45	81.60
Alirezazadeh *et al.* [3] (2015)	/	/	81.30	86.15
Zhou *et al.* [197] (2016)	/	/	78.60	75.70
Wu *et al.* [168] (2016)	/	81.19	/	/
Yan [172] (2017)	/	/	66.30	78.70
Lu *et al.* [101] (2017)	/	84.15	**83.50**	84.30
Proposed-Gray	**77.60**	83.63	77.59	78.05
Proposed-Color	**81.50**	**88.11**	80.00	**87.60**

3.4 Conclusion

In this chapter, we proposed an approach to the problem of kinship verification through face texture representations in components of various color spaces. We experimented with three texture descriptors (MSLPQ, MSBSIF and MSCoALBP) and their fusion in seven color spaces (RGB, HSL, HSV, Lab, Luv, YCbCr and YUV), considering the gray-scale as baseline. Thorough experiments are performed on the available kinship facial databases, namely the Cornell KinFace, TSKinFace and KinFaceW. The obtained results show the effectiveness of using the color information for kinship verification compared with the gray-scale counterparts. Generaly, HSV color space outperforms the other spaces. These results point out the importance of face color information for kinship verification in all scenarios. Additionally, our results compare favorably against the recent approaches in the literature on the benchmark databases. As a future work fusion of different color spaces will be investigated. Also, generate new color space by using mixture of different color spaces channels (chrominance/luminance) for kinship verification.

4 Learning Multi-view Deep and Shallow Features through new Discriminative Subspace for Bi-subject and Tri-subject Kinship Verification

Contents



4.1 Introduction

The learning of metric distance or differential function of data plays an important role in various visual analysis tasks such as multimedia retrieval [167], image clustering [171] and object recognition [159]. While a number of distance metric learning methods have been proposed over the past decade [31,47,53,116,143,159,171]. In general, these metric learning methods mainly focus on learning a single distance scale, either in a monolithic pattern to represent the feature, or sequentially representing multiple types of features, which are usually deficient: first, complementary information cannot be used well of multiview representations; another is that such a direct sequential manner ignores the importance of different perspectives and a short chain of meaningful explanation physically because each single viewing feature possesses its own characteristics. While many multi-view distance learning methods have been introduced [60, 101, 103, 174, 193], these methods generally assign multiple data terrain in an area so latent that they cannot maintain a property specific to each view.

In this chapter, we present a deep and shallow multi-view metric learning (SILD+WCCN/LR) method by exploiting more traits of multi-view data. The SILD+WCCN/LR jointly learns an optimal combination of multiple scores on multi-view representations, in which it not only learns an individual distance metric for each view to retain its specific property but also learns a shared representation for different views in a unified latent subspace to preserve their common properties. In addition, to exploit the nonlinear structure of data points, we used deep and shallow features (multi-view) through the proposed SILD+WCCN combined with Logistic Regression (LR) [56] method (score fusion strategy). VGG-Face [128] method shows best performances for face verification topic, and obtained 98.95% on LFW database. Inspired by this work, we used in our work deep and shallow features projected into the discriminative subspace of the proposed SILD+WCCN method combined with LR method for kinship verification. In our frameworks, the set of face images are represented as a deep and shallow features of the different descriptors, VGG-Face [128], Multi-scale Local Phase Quantization (MSLPQ) [120], Multi-scale local Binarised Statistical Image Features (MSBSIF) [77], Multi-Scale Local Binary Patterns (MSLBP) [119], the Histogram of Oriented Gradients (HOG) [29] and Multi-Scale Co-occurrence of Adjacent Local Binary Patterns (CoALBP) [117]. The contributions of this work are summarized as follows:

1. We propose a new discriminative subspace of the proposed Side-Information based Linear Discriminant analysis integrating Within Class Covariance Normalization (SILD+WCCN) subspace transformation method for kinship verification. Thus, WCCN decreases the class intra-variability effect by reducing the expected classification error on the training step [9].

2. We propose a two robust automated kinship verification frameworks suitable for

bi-subject and tri-subject kinship verification, from face images captured in unconstrained environments. The face data is represented as a multi-view feature based on the combination of different deep and shallow features in order to provide a more powerful face model.

3. We investigate the efficiency of deep/shallow information through new discriminative subspace using the two-step learning approach, SILD+WCCN, and Logistic Regression, for automatic verification of kinship from facial images.

4. We extensively evaluate our SILD+WCCN/LR method against the state-of-the-art methods using two challenging kinship databases namely KinFaceW-II and TSKinFace.

The chapter is organized as follows: Our bi-subject and tri-subject kinship verification frameworks are presented in Section 4.2. The experimental data and setup are presented and results are given in Section 4.3. Finally, concluding remarks are given in Section 4.4.

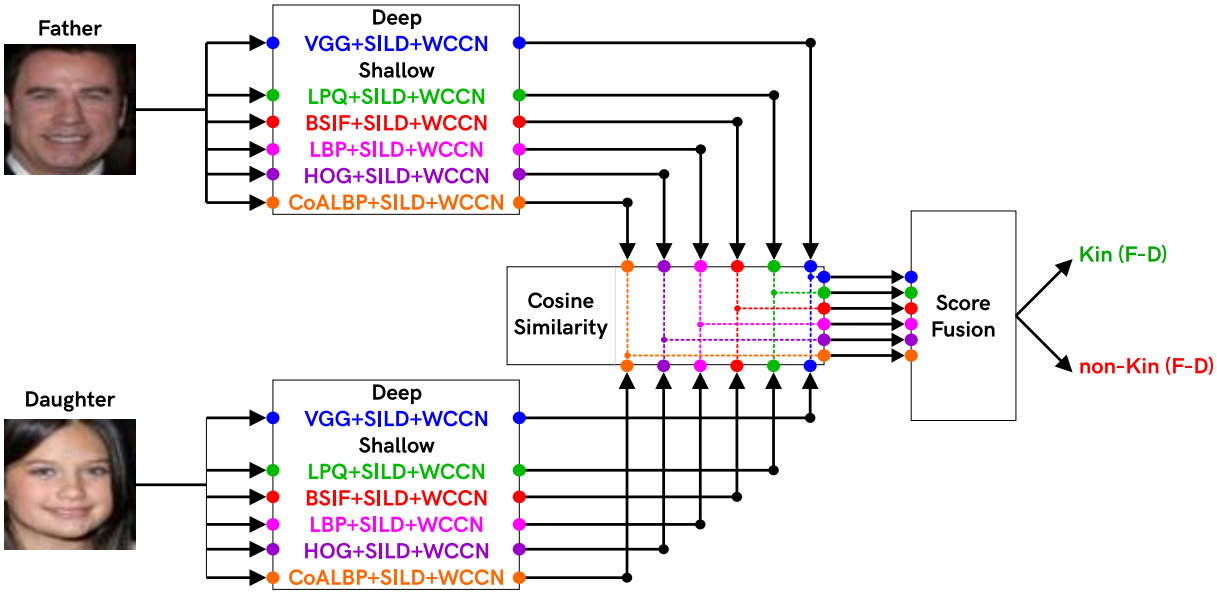
4.2 Proposed kinship verification frameworks

We aim to investigate the effectiveness of using deep and shallow features through the proposed SILD+WCCN method. For this issue, we propose two frameworks for Bi-Subject matching and Tri-Subject matching problems. Figure 4.1 shows our two frameworks scheme. The Parent-Child pair of face images is given as an input. First, the pair face images is cropped and normalized into an $X \times Y$ pixels. Then, we extract the features from each face image (i.e. deep and shallow features) using five layers of VGG-Face [128] (i.e. *pool5, fc6, relu6, fc7* and *relu7*) and five local texture/shape descriptors (i.e. *LPQ, BSIF, LBP, HOG* and *CoALBP*). The features are then projected into the proposed (SILD+WCCN) subspace.

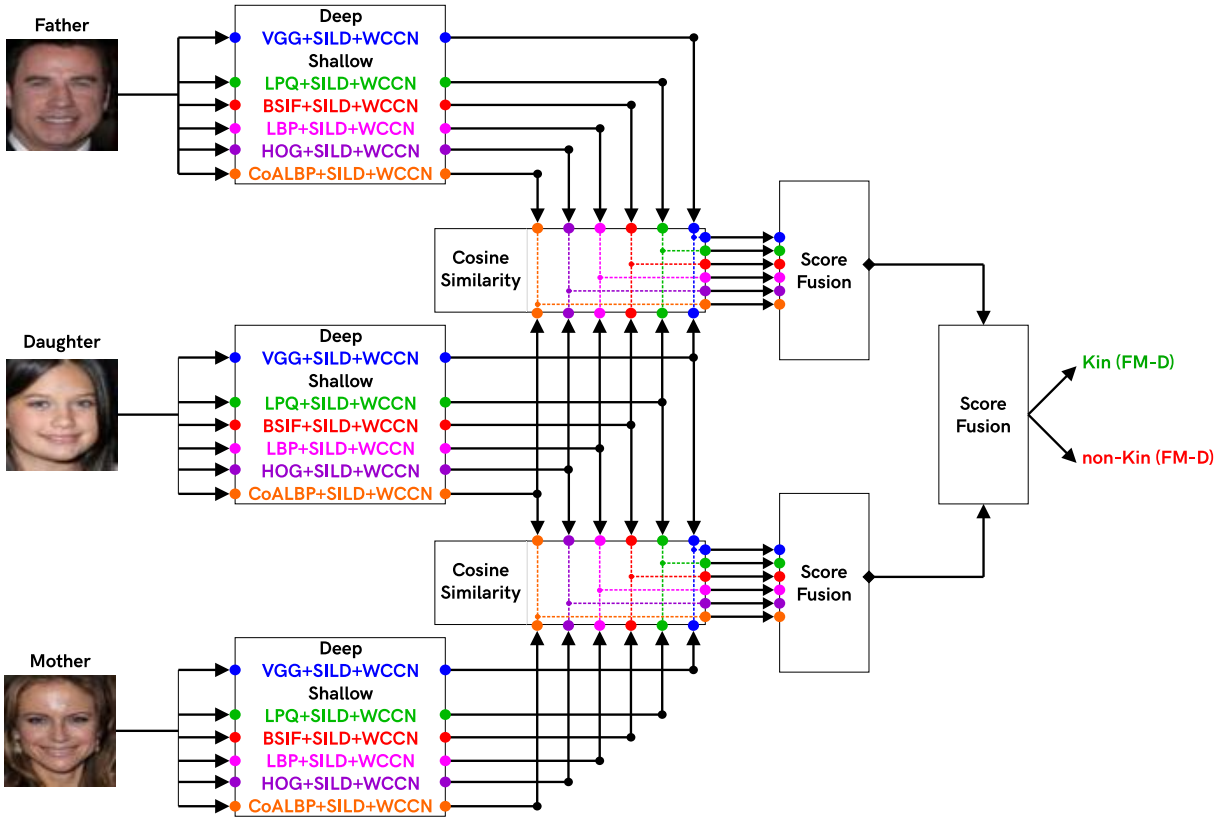
Bi-subject matching: We compute the cosine similarity between the projected feature vectors of one pair for each feature type (Deep&Shallow). We apply score fusion of the six scores resulting from the considered pair face image using Logistic Regression method (LR) [56]. Finally, the fusion score is compared to a threshold, set from the Receiver Operating Characteristic (ROC) curve during performance evaluation, to decide whether the pair belongs to the persons from the same family or not.

Tri-subject matching: We compute the cosine similarity between the projected feature vectors of the two pairs (F-S and M-S for FM-S tri-Subject, and F-D and M-D for FM-D tri-Subject) for each feature type (Deep&Shallow). We refer that F, M, S, and D mean, Father, Mother, Son, and Daughter, respectively. We apply score fusion for two pairs of the six scores resulting from the first pair and the six scores resulting from the second pair face image using Logistic Regression method [56]. Finally, the fusion score is compared to a

threshold, set from the Receiver Operating Characteristic (ROC) curve during performance evaluation, to decide whether the Child belongs to the persons from the same family (Father and Mother) or not. In the following we provide the details of the steps of our approach.



(a)



(b)

Figure 4.1: The two proposed kinship verification frameworks using deep and shallow features through the proposed SILD+WCCN method, (a) Bi-Subject framework and (b) Tri-Subject framework.

4.2.1 Framework design

In the offline (training) stage, the optimal projection matrices are estimated, and in online (test) stage, new samples are projected by these matrices and matched. The training is constructed using deep and shallow features extracted from the training face images. The six models of the proposed SILD+WCCN method are defined as follows: VGG+SILD+WCCN, LPQ+SILD+WCCN, BSIF+SILD+WCCN, LBP+SILD+WCCN, HOG+SILD+WCCN and CoALBP+SILD+WCCN.

The training data of SILD+WCCN method includes the positive pairs and negative pairs. The positive pairs are used to compute the covariance matrix of the within-class scatter matrix and the negative pairs are used to compute the covariance matrix of the between-class scatter matrix of SILD+WCCN method.

In the test phase, each face pair passes the same steps of feature extraction as in the training phase, then each pair projected, in the dimensionality reduction and classification (SILD+WCCN). Finally, the cosine similarity is used to check whether the pair of reduced features are match (belonging to the same family) or not.

4.2.2 Feature extraction

To describe face images, we extracted two kind of features (Deep and Shallow):

Deep features (VGG-Face [128]): We extracted deep features from color face images of size $224 \times 224 \times 3$ using five layers: the **pool5 layer** consists of $7 \times 7 \times 512$ weights (25,088D). The **fc6 layer** is the sixth fully connected layer which contains 4,096D weights. The **relu6 layer** is the sixth Rectified Linear Unit (ReLU) activation function layer which contains 4,096D weights. The **fc7 layer** is the seventh fully connected layer which contains 4,096D weights. The **relu7 layer** is the seventh Rectified Linear Unit (ReLU) activation function layer which contains 4,096D weights. We concatenate all weights resulted from the five layers of VGG method.

Shallow features (texture/shape features): Regarding shallow features extraction, we resize the gray face images into $64 \times 64 \times 1$. Five local texture/shape descriptors (i.e. LPQ, BSIF, LBP, HOG and CoALBP) was used. We use eight filters with different sizes $W = \{3, 5, 7\}$ in the Multi-Scale Binarized Statistical Image Features (MSBSIF) [77]. In the Multi-Scale Local Phase Quantization (LPQ) [120], the window size is $M = \{3, 5, 7\}$. The choice of these scales is to depend on the size of the face image. We extract at Multi-Scale the Local Binary Patterns (LBP) [119], the radius $R = \{1, 2, 3\}$ and the number of pixels in the neighborhood $P = 8$. Every face image is subdivided into 64 blocks, each of size 8×8 pixels. We use histograms of 256 bins to aggregate the local features extracted from each block. By concatenating the histograms of all the 64 blocks, a vector of dimension $3 \times 64 \times 256$ is obtained. For Histogram of Oriented Gradients (HOG) [29], the window size is 8×8. For Multi-Scale Co-occurrence of Adjacent Local Binary Patterns

(CoALBP) [117], the extracted features from each face image are the combination of three histograms where the size is 3×1024, using the LBP+ operator with radius R = {1, 2, 4} and the identical directions specified by the distances S = {2, 4, 8}.

4.2.3 Side-Information based Linear Discriminant analysis (SILD)

Supervised subspace transformation methods, such as Linear Discriminant Analysis (LDA) [127] and Exponential Discriminant Analysis (EDA) [190], enhance the discrimination of the extracted features by transform these data into subspace where make easier their separation and classification. These methods need the handy of class information for each sample where the classes are supervised of each sample, and the within class scatter matrix (S_w) and the between class scatter matrix (S_b) must be calculated with full label information.

That means, LDA and EDA fail in weakly labeled data. Kan et al. [109] proposed a new representation to resolve this problem by directly operating the S_w and S_b matrices with the side-information. While, the positive classes pair images are directly utilized to calculate the within class scatter matrix and the negative classes pair images are used to compute the between class scatter matrix. Let us refer that $P_{class} = \left\{(\check{\xi}_i^1, \hat{\xi}_i^1) : l(\check{\xi}_i^1) = l(\hat{\xi}_i^1)\right\}$ as the collection of positive-class image pairs and $N_{class} = \left\{(\check{\xi}_i^0, \hat{\xi}_i^0) : l(\check{\xi}_i^0) \neq l(\hat{\xi}_i^0)\right\}$ as the collection of negative-class image pairs, where the image ξ is represented by the class label $l(\xi)$. Here, the within-class and between-class scatter matrices of Side-Information based Linear Discriminant analysis (SILD) method can be represented by:

$$S_w^{sild} = \sum_{i=1}^{C_1} (\check{\xi}_i^1 - \hat{\xi}_i^1)(\check{\xi}_i^1 - \hat{\xi}_i^1)^T \tag{4.1}$$

$$S_b^{sild} = \sum_{i=1}^{C_0} (\check{\xi}_i^0 - \hat{\xi}_i^0)(\check{\xi}_i^0 - \hat{\xi}_i^0)^T \tag{4.2}$$

The target function for SILD is:

$$U_{opt}^{sild} = argmax_U \frac{U^T S_b^{sild} U}{U^T S_w^{sild} U} \tag{4.3}$$

The problem in 4.3 can be solved by a two-step method [145]. Firstly, S_w is diagonalized as follows:

$$S_w = H \Lambda H^T \tag{4.4}$$

$$(H \Lambda^{-1/2})^T S_w (H \Lambda^{-1/2}) = I \tag{4.5}$$

Secondly, S_b is also diagonalized:

$$(H\Lambda^{-1/2})^T S_b (H\Lambda^{-1/2}) = ZEZ^T \tag{4.6}$$

Finally, the projection matrix can be computed as:

$$U^{sild} = H\Lambda^{-1/2} Z \tag{4.7}$$

where H and Z are orthogonal matrices and Λ and E are diagonal matrices.

A solution to the optimization problem in (4.3) is obtained via solving the generalized eigenvalue problem. The projection matrix of SILD is formed by the first k eigenvectors in (4.7) that ordered in the descending order of eigenvalues.

In Eq. (4.3), the inter-class distance of training samples from different subjects $(\check{\xi}_i^0, \hat{\xi}_i^0)|_{i=0}^{C_0}$ for all pairs is maximized by the numerator while the intra-class distance of training samples from the same subjects $(\check{\xi}_i^1, \hat{\xi}_i^1)|_{i=0}^{C_1}$ for all pairs is minimized by the denominator. This equation derived from the following multiobjective programming problem:

$$\begin{gathered} max \frac{U^T S_b^{sild} U}{U^T U} \\ min \frac{U^T S_w^{sild} U}{U^T U} \end{gathered} \tag{4.8}$$

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

4.2.4 Within-class covariance normalization

The first use of the Within-Class Covariance Normalization (WCCN) is in the community of speaker recognition. While Dehak *et al.* [34] founded that it is the best technique to project the reduced-vectors of LDA method to a new subspace determined by the square-root of the inverse of the within-class covariance matrix. We propose a new variant of SILD by integrating WCCN:

$$W = \sum_{i=1}^{C_1} \frac{(U^{sild})^T \check{\xi}_i^1 - (U^{sild})^T \hat{\xi}_i^1}{(U^{sild})^T \check{\xi}_i^1 - (U^{sild})^T \hat{\xi}_i^1} \tag{4.9}$$

where, U^{sild} is the SILD projection matrix found in Eq.4.7. The WCCN projection matrix C is obtained by Cholesky decomposition [68,184] of the inverse of W: $W^{-1} = CC^T$. Where the new projection matrix $B^{sild+wccn}$ is obtained by: $B^{sild+wccn} = C^T U^{sild}$. By imposing upper bounds on the classification error metric [9], WCCN decreases the within-class variations effect by reducing the expected classification error on the training step.

The procedure of this proposed method, SILD+WCCN, is detailed in algorithm 1.

Algorithm 1 SILD+WCCN

Input:
- The matrix ξ of the N training samples.
- The weak labels ($labels_W$) for extracting the positive-class image pairs $P_{class} = \left\{(\check{\xi}_i^1, \hat{\xi}_i^1) : l(\check{\xi}_i^1) = l(\hat{\xi}_i^1)\right\}$ and negative-class image pairs $N_{class} = \left\{(\check{\xi}_i^0, \hat{\xi}_i^0) : l(\check{\xi}_i^0) \neq l(\hat{\xi}_i^0)\right\}$

Output:
- The projection matrix $B^{sild+wccn}$ of SILD+WCCN.

1: $S_w^{sild} = \sum_{i=1}^{C_1}(\check{\xi}_i^1 - \hat{\xi}_i^1)(\check{\xi}_i^1 - \hat{\xi}_i^1)^T$
2: $S_b^{sild} = \sum_{i=1}^{C_0}(\check{\xi}_i^0 - \hat{\xi}_i^0)(\check{\xi}_i^0 - \hat{\xi}_i^0)^T$
3: Compute the matrices: S_w^{sild} and S_b^{sild}
4: Sort the m eigenvectors $U^{sild} = (HZ)_k$ according to $\Lambda_k^{-1/2}$ in decreasing order.
5: $W = \sum_{i=1}^{C_1}((U^{sild})^T\check{\xi}_i^1 - (U^{sild})^T\hat{\xi}_i^1)((U^{sild})^T\check{\xi}_i^1 - (U^{sild})^T\hat{\xi}_i^1)^T$
6: Compute WCCN projection matrix (C): $W^{-1} = CC^T$
7: Compute the new $B^{sild+wccn} = C^T U^{sild}$

4.2.5 Matching

To compare between two faces pair, we used the six reduced deep and shallow features projected through the proposed SILD+WCCN subspace. Then, we applied cosine similarity [116] for each of the six features from pair test of the two face images, so a six match scores are done. After discriminant analysis methods, the using of cosine similarity distance has an advantage comes from its connection to the Bayes decision rule [95]. Cosine similarity between two vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as the following:

$$cos(\mathbf{X}, \mathbf{Y}) = \frac{\mathbf{X}^T\mathbf{Y}}{\|\mathbf{X}\| \, \|\mathbf{Y}\|} \tag{4.10}$$

Where $\|.\|$ is the Euclidean norm. A high value of the produced score means a high probability that $\mathbf{X}$ and $\mathbf{Y}$ are same family.

4.2.6 Score Fusion

Score fusion interested in normalizing the multiple scores and then fusing them. The similarity scores, for deep feature, and the loglikelihood scores, for shallow scores, were normalized to produce probabilities such that the scores lie in the range [0, 1]. We got this by using logistic regression [56] method, hence taking y_i as the input and generating probability, P_i, as the output:

$$P_i = (1 + exp(ay_i + b))^{-1} \tag{4.11}$$

where a is a scaling factor and b is a bias.

We chose to use logistic regression for two reasons. First, it is a well establish method in

statistics, pertinent to a family of methods called generalized linear models. Its optimization procedure, known as "gradient ascent" [57], is well understood; it converges to a global minimum, meaning that there is a unique solution. Second, its output can be interpreted as a probability and so presenting this information to the user is more meaningful than just the raw score.

The final combined score is obtained by taking the product of the two scores, $P_{Feature1}$ for first feature and $P_{Feature2}$ for second feature. Although the sum rule could have also been used. Both rules result in similar generalization performance.

4.3 Experiments

In this section, we perform a number of experiments to evaluate the proposed kinship verification systems and investigate the strengths of the proposed SILD+WCCN. Firstly, we present the benchmark databases used in our experiments. Then, we discuss the parameter settings used for kinship verification. Finally, we provide and discuss the results and compare them with those of the state of the art.

4.3.1 Parameter Settings

Our approach's performance is evaluated based on the same experimental protocol found in the literature works [103, 172, 174], in which five-fold cross-validation for kin verification is performed by keeping the same number of pairs images for each fold. This protocol is followed to make sure that our results are directly comparable to the state of the art. The negative pairs for kinship are generated randomly such that each image appears only once in the training set. The number of positive pairs and negative pairs is the same for both training and test. For score fusion, we use LR method and two SVMs methods [126] (i.e. linear SVM and Kernel SVM). For performance evaluation, we use 4 folds for training and the remaining fold for testing through five-folds cross-validation for both LR and linear SVM methods. For Kernel SVM (with RBF kernel), we use 3 folds for training, one fold for performance evaluation (parameter tuning) of RBF kernel and the last fold for testing through five-folds cross-validation. For the SVM methods, the code we developed is based on OSU SVM Classifier Matlab Toolbox [104] which implements a Matlab interface to access LIBSVM [23]. For linear SVM, the value of penalty parameter is $C = 1$. For kernel SVM, the penalty parameter $C \in [0, 1000]$ and the standard deviation of RBF kernel $\sigma \in [0.01, 10]$.

4.3.2 Results and discussion

In this subsection, we introduce and discuss the evaluation's results of the proposed approach. Moreover, all the experiments were done for the original approach (SILD

method), which works as a baseline for evaluating the proposed SILD+WCCN method. Furthermore, the performances of the deep and shallow features are separately examined as well as their fusion.

Experiments on LFW [67]: To show the effectiveness of the proposed method, we compared our SILD+WCCN method to SILD [109], PHL+SILD [76], DDMML [101] and MvDML [60] methods for face verification on LFW database as shown in Table 4.1. In this comparison, we compared our method with the methods that used only shallow features under image-restricted setting. Our SILD+WCCN method used MSBSIF as a baseline descriptor.

Table 4.1: Performances (mean accuracy ± standard error (%)) of our SILD+WCCN compared with state of the art on LFW database using only shallow features under image-restricted setting.

Method	Mean Accuracy ± Standard Error (%)
SILD [109]	87.68 ± 0.50
PHL+SILD [76]	88.67 ± 0.70
DDMML [101]	93.28 ± 0.39
MvDML [60]	93.27 ± 0.28
MSBSIF+SILD (Our)	86.17 ± 1.05
MSBSIF+SILD+WCCN (Our)	**94.63 ± 0.86**

From this comparison, we see that our method got the first rank by using only MSBSIF features description. Unlike to SILD, PHL+SILD, DDMML and MvDML methods which used the combination of various features. SILD and PHL+SILD methods used eight features using Original and Square root of Intensity, LBP, Gabor and Block Gabor. PHL+SILD improves SILD with 1% using the same features description (i.e. the combination of eight features), unlike to our SILD+WCCN method which improves SILD method with 8% using the same features (i.e. MSBSIF description). This demonstrates that SILD method fails to project MSBSIF features into discriminative subspace, unlike our SILD+WCCN which projects these features to get more better discrimination.

This first experiment suggests that MSBSIF features is more discriminative through the proposed subspace transformation method (i.e. SILD+WCCN) than SILD for face verification from face images. The corresponding ROC curves of our SILD+WCCN and SILD methods on LFW database are depicted in Figure 4.2.

Face recognition using deep features (i.e. VGG-Face [128]) shows interesting and motivating results (i.e. 98.95% on LFW database). We inspired by the work of Parkhi et al. [128] to project VGG-Face features into the discriminative subspace of the proposed SILD+WCCN method for kinship verification.

Experiments on KinFaceW-II [103] and TSKinFace [133]: Tables 4.2 and 4.3 show kinship verification accuracy from different deep and shallow features and their fusion using the proposed SILD+WCCN/LR method compared with SILD/SVM, SILD/KSVM, SILD/LR, SILD+WCCN/SVM and SILD+WCCN/KSVM methods on KinFaceW-II and

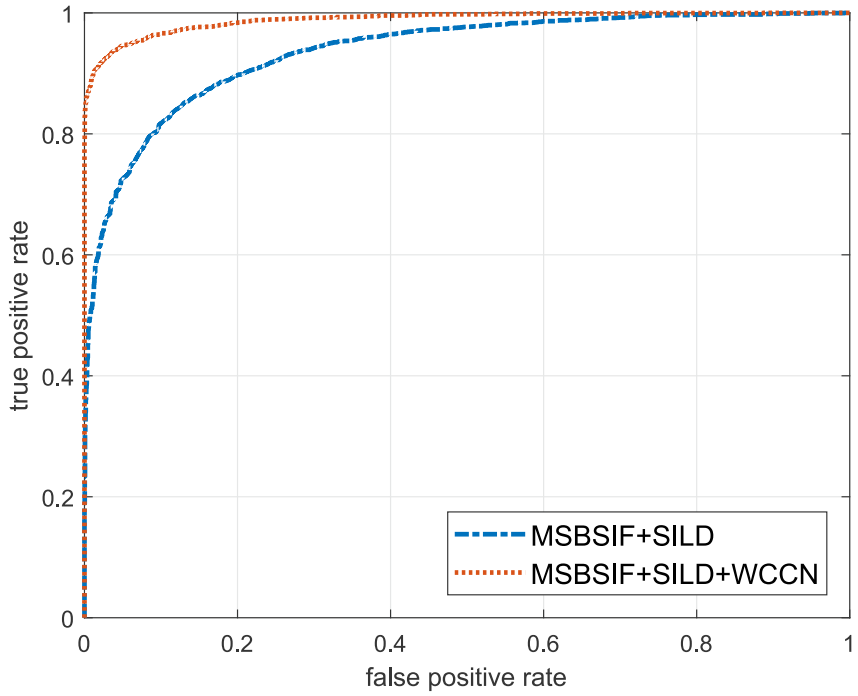


Figure 4.2: ROC curves of MSBSIF+SILD+WCCN and MSBSIF+SILD methods on the LFW database.

TSKinFace databases, respectively. These tables show that our SILD+WCCN method improves the SILD method with a large margin. The performance is improved with variation between 3% and 6%. It is also noticeable that the best results are obtained by (Deep&Shallow) features description.

We start by studying the kinship verification accuracy using six features description separately. These results, given in Tables 4.2 and 4.3. The combinations of deep and shallow features used for multi-view kinship verification. We also used the VGG descriptor that is based on a deep network [128], and was thus trained using a training set other than the KinFaceW-II and TSKinFace databases. It follows that the proposed multiview-based SILD+WCCN/LR approach outperforms the other approaches for all features description separately. In particular, when using the VGG, MSLPQ and MSBISF features, our SILD+WCCN method shows the highest kinship verification accuracy. We compare the accuracy of multi-view deep and shallow features for kinship verification based on the SILD+WCCN/LR approach on Father-Son (F- S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D) relations. For that we considered six multi-view of deep and shallow features that were shown to have the highest kinship verification accuracy in Tables 4.2 and 4.3.

Table 4.2: Performance comparisons (%) with the metric learning methods on KinFaceW-II database.

Method	Feature	F-S	F-D	M-S	M-D	Mean
SILD	LPQ	78.20	74.60	77.00	72.60	75.60
SILD	BSIF	80.40	75.20	76.60	73.00	76.30
SILD	LBP	79.00	72.80	72.60	70.20	73.65
SILD	HOG	73.00	66.00	69.80	65.20	68.50
SILD	CoALBP	71.00	71.00	68.40	69.00	69.85
SILD	VGG	75.40	70.60	75.00	75.80	74.20
SILD/SVM	Deep&Shallow	82.40	77.40	79.20	77.40	79.10
SILD/KSVM	Deep&Shallow	82.00	77.60	79.40	78.40	79.35
SILD/LR	Deep&Shallow	85.20	77.40	81.20	79.00	80.70
SILD+WCCN	LPQ	84.20	80.60	78.00	77.40	80.05
SILD+WCCN	BSIF	83.80	77.20	79.80	78.80	79.90
SILD+WCCN	LBP	82.80	79.20	74.80	73.20	77.50
SILD+WCCN	HOG	72.80	74.80	72.80	72.40	73.20
SILD+WCCN	CoALBP	76.80	75.60	76.20	74.80	75.85
SILD+WCCN	VGG	78.20	74.00	76.80	80.20	77.30
SILD+WCCN/SVM	Deep&Shallow	86.20	82.60	84.20	83.80	84.20
SILD+WCCN/KSVM	Deep&Shallow	85.80	82.60	83.80	83.40	83.90
SILD+WCCN/LR	Deep&Shallow	**88.40**	**84.20**	**85.80**	**86.40**	**86.20**

Table 4.3: Performance comparisons (%) with the metric learning methods on TSKinFace database.

Method	Feature	F-S	F-D	M-S	M-D	Mean	FM-S	FM-D
SILD	LPQ	81.09	79.37	80.60	79.28	80.09	84.12	83.06
SILD	BSIF	80.59	80.36	81.20	80.88	80.76	83.92	83.05
SILD	LBP	77.18	76.89	78.54	76.71	77.33	81.37	79.19
SILD	HOG	73.38	72.30	73.78	73.90	73.34	76.52	75.49
SILD	CoALBP	71.32	73.60	70.76	71.42	71.78	73.19	74.50
SILD	VGG	72.13	73.00	73.40	77.29	73.96	76.32	78.77
SILD/SVM	Deep&Shallow	82.06	81.46	82.96	82.17	82.16	84.99	84.65
SILD/KSVM	Deep&Shallow	82.06	81.45	82.46	82.13	82.03	84.78	84.71
SILD/LR	Deep&Shallow	81.66	81.96	82.65	82.87	82.29	85.38	85.05
SILD+WCCN	LPQ	86.64	85.06	85.30	85.26	85.57	87.82	88.15
SILD+WCCN	BSIF	85.76	84.66	83.55	86.25	85.06	86.75	87.15
SILD+WCCN	LBP	80.21	79.37	80.32	81.96	80.47	84.12	83.95
SILD+WCCN	HOG	76.71	73.51	77.29	76.99	76.13	79.24	78.49
SILD+WCCN	CoALBP	79.04	77.59	80.20	78.79	78.91	82.35	79.58
SILD+WCCN	VGG	84.12	82.68	85.87	85.75	84.61	87.63	87.84
SILD+WCCN/SVM	Deep&Shallow	88.40	**87.15**	88.11	89.54	88.30	90.36	90.54
SILD+WCCN/KSVM	Deep&Shallow	88.16	87.00	87.69	89.58	88.11	90.50	90.97
SILD+WCCN/LR	Deep&Shallow	**89.08**	87.05	**88.59**	**89.63**	**88.59**	**90.94**	**91.23**

For each specific case of Tables 4.2 and 4.3 (regardless the features, databases, and relations types), the proposed approach SILD+WCCN performs better than the SILD counterpart which obviously demonstrates the effectiveness of the proposed SILD+WCCN method. Moreover, our results demonstrate that SILD+WCCN is able to extract better discriminative features than SILD method. Figures 4.3 and 4.4 show the performance

comparisons of projection the different deep and shallow features (i.e. LPQ, BSIF, LBP, HOG, CoALBP, VGG and Deep&Shallow) using the proposed SILD+WCCN method compared with SILD method on KinFaceW-II and TSKinFace databases, respectively. In these figures, the abbreviations D&S/SVM, D&S/KSVM and D&S/LR represent the Deep&Shallow features fusion using the three different score fusion methodology. These figures confirm the previous remarks and clearly depict that the proposed SILD+WCCN greatly outperforms the SILD method using the different features description for both KinFaceW-II and TSKinFace databases. Furthermore, all the score fusion methodology (i.e. linear SVM, kernel SVM and LR) extract the complementarity of the different deep and shallow features and each of these methods shows performances better than each feature's description individually. Moreover, LR method performs better than linear SVM and kernel SVM methods in the most cases.

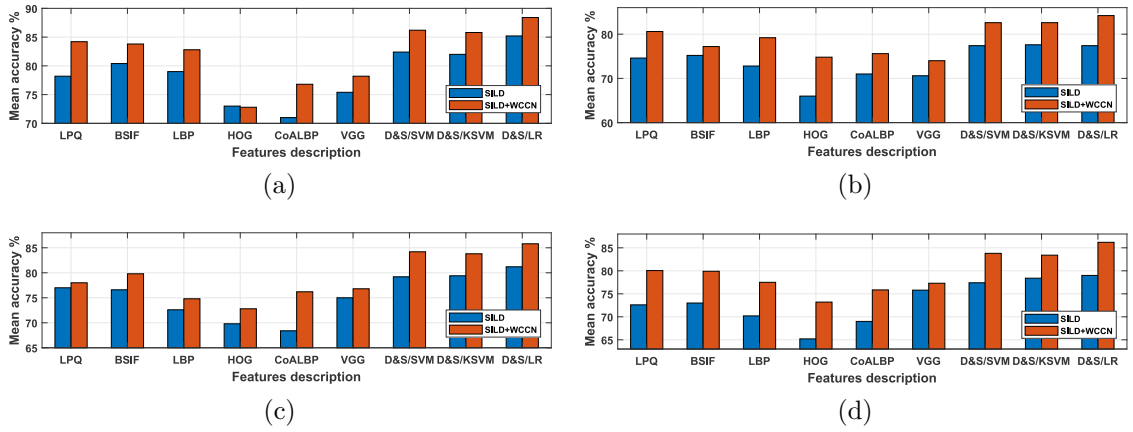


Figure 4.3: Performance comparisons (%) of projection the different deep and shallow features on KinFaceW-II database using the proposed SILD+WCCN method compared with SILD method, obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

To check the performance of different kinship relations, we plot in Figures 4.5 and 4.6 the ROC curves of different methods (SILD+WCCN and SILD) using different deep and shallow features on KinFaceW-II and TSKinFace databases. For clarity purpose, we plot the deep and shallow features separately as well as their fusion (Deep&Shallow). The ROC curves confirm the previous remarks for individual relations. Deep&Shallow features are better than each feature individually and SILD performances are less than the proposed SILD+WCCN metric learning method.

To analyze the effect of SILD/LR method and the proposed SILD+WCCN/LR method on kinship verification, we perform an experiment on the face similarities. For a general analysis, we plot, in Figure 4.7, the distributions of the similarities of positive pairs and negative pairs for the all four relations types on KinFaceW-II and TSKinFace databases. The similarities between face pairs in different methods are computed using Deep&Shallow feature description. The positive and negative pairs similarities in SILD are dense and their overlap is considerable while the distributions in SILD+WCCN subspace is spread and the

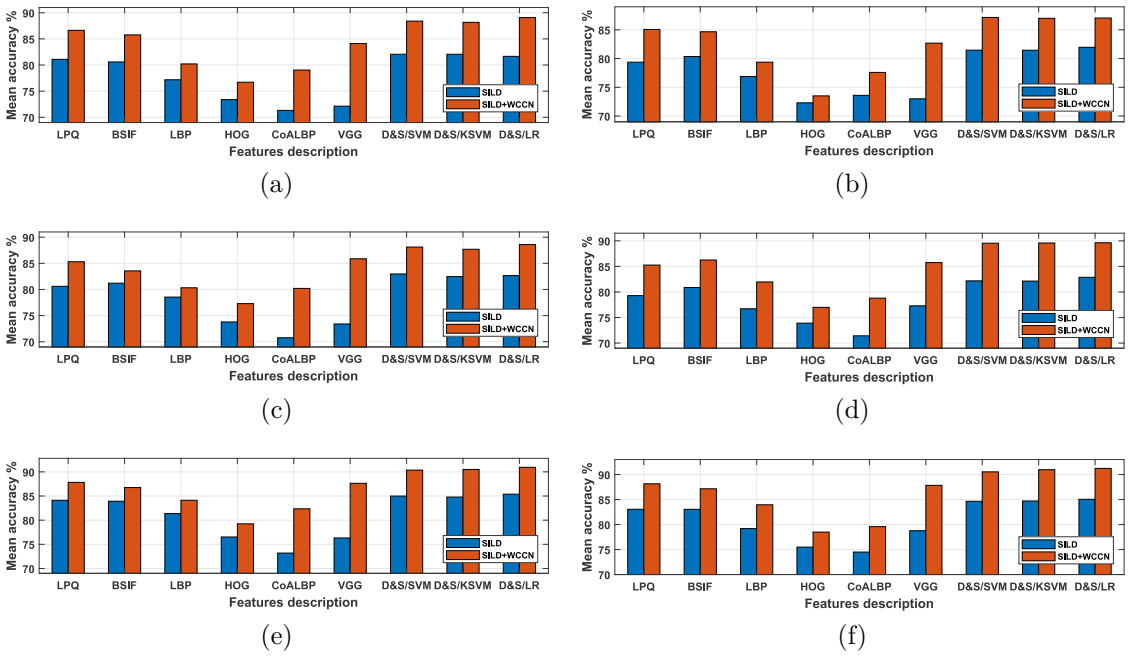


Figure 4.4: Performance comparisons (%) of projection the different deep and shallow features on TSKinFace database using the proposed SILD+WCCN method compared with SILD method, obtained on (a) F-S set, (b) F-D set, (c) M-S set, (d) M-D set, (e) FM-S set and (f) FM-D set, respectively.

overlap is less significant. This experiment suggests that deep and shallow features are more discriminative through the proposed subspace transformation method (i.e. SILD+WCCN) for kinship verification from face images.

The corresponding ROC curves of different methods (SILD and SILD+WCCN) using the different deep and shallow features on TSKinFace database for FMS and FMD relations are depicted in Figure 4.8. This figure clearly depicts that our SILD+WCCN metric learning method projects the face images pairs to discriminative subspace which facilitate the score learning (score fusion) with the score fusion methods. Furthermore, our tri-subject framework performs better than the bi-subject framework, benefiting the core of family on Father-Mother-Son and Father-Mother-Daughter relations. Where the tri-subject framework using FMS relation face images performs better than the bi-subject framework using the Father-Son and Mother-Son separately relations, and the tri-subject framework using FMD relation face images performs better than the bi-subject framework using the Father-Daughter and Mother-Daughter separately relations.

4.3.3 Comparison with the results of the state of the art

The best kinship verification performances of our approach are achieved using deep and shallow features (Deep&Shallow) on KinFaceW-II and TSKinFace databases. Our SILD+WCCN/LR approach based on multi-view deep and shallow features outperforms the current state of the art multi-view metric learning including [60, 101, 103, 174, 175, 193] on KinFaceW-II and TSKinFace databases for bi-subject and tri-subject matching problems.

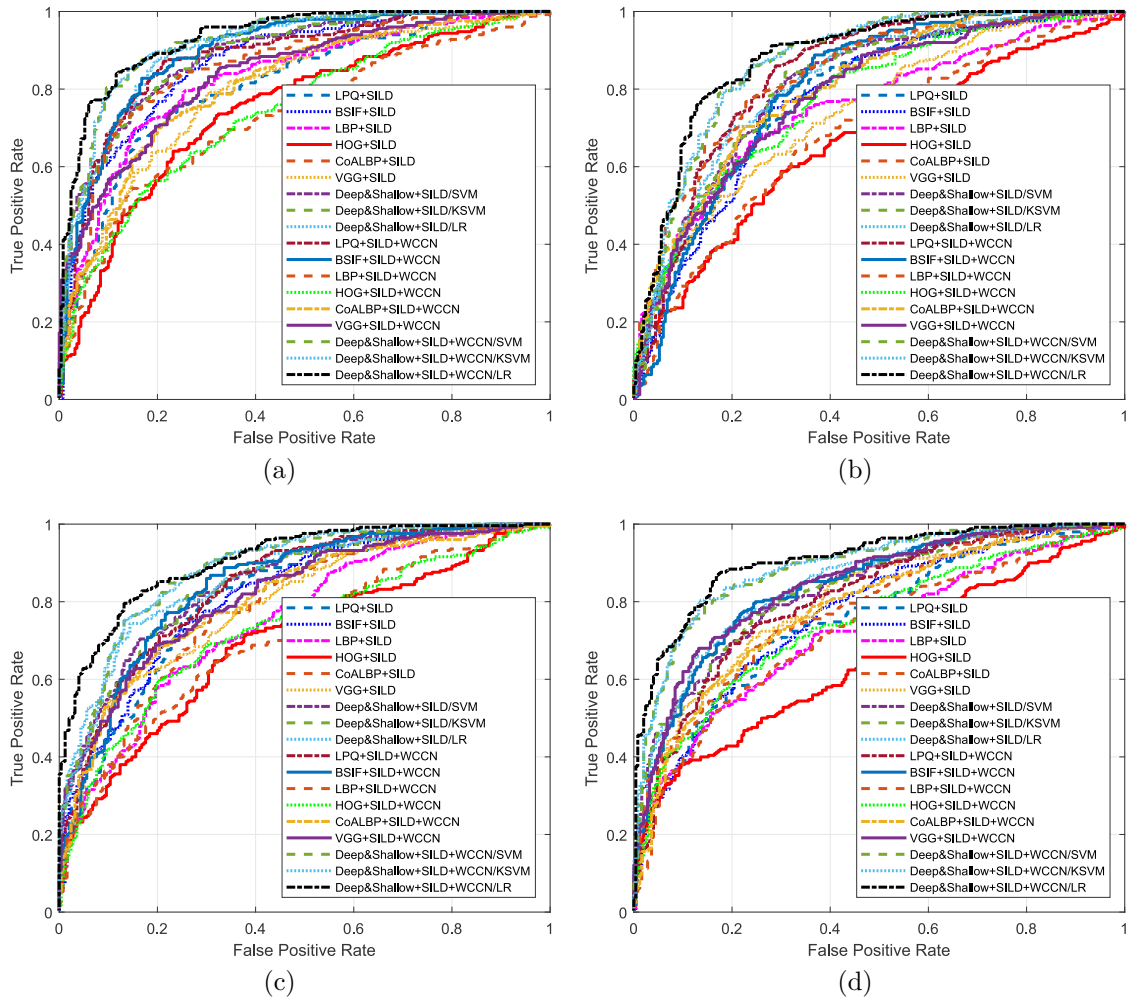


Figure 4.5: ROC curves of different methods (SILD and SILD+WCCN) using the different deep and shallow features on KinFaceW-II database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

Bi-subject face matching: For the SILD/LR method, verification rates of 80.70% and 82.29% are reported on KinFaceW-II and TSKinFace databases, respectively. For SILD+WCCN/LR method, verification rates of 86.20% and 88.59% are reported on KinFaceW-II and TSKinFace databases, respectively. These results are compared with the state of the art in Tables 4.4 and 4.5. The comparison reveals that the proposed SILD+WCCN/LR method outperforms the recent state of the art on the two databases, KinFaceW-II and TSKinFace.

Tri-subject face matching: For the SILD/LR method, verification rates of 85.38% and 85.05% are reported on TSKinFace database for FM-S and FM-D relations, respectively. For SILD+WCCN/LR method, verification rates of 90.94% and 91.23% are reported on TSKinFace database for FM-S and FM-D relations, respectively. These results are compared with the state of the art in Table 4.5. The comparison reveals that the proposed SILD+WCCN/LR method outperforms the recent state of the art on the TSKinFace database for both FM-S and FM-D relations.

Proposed SILD+WCCN method vs. SILD method: From the evaluation on

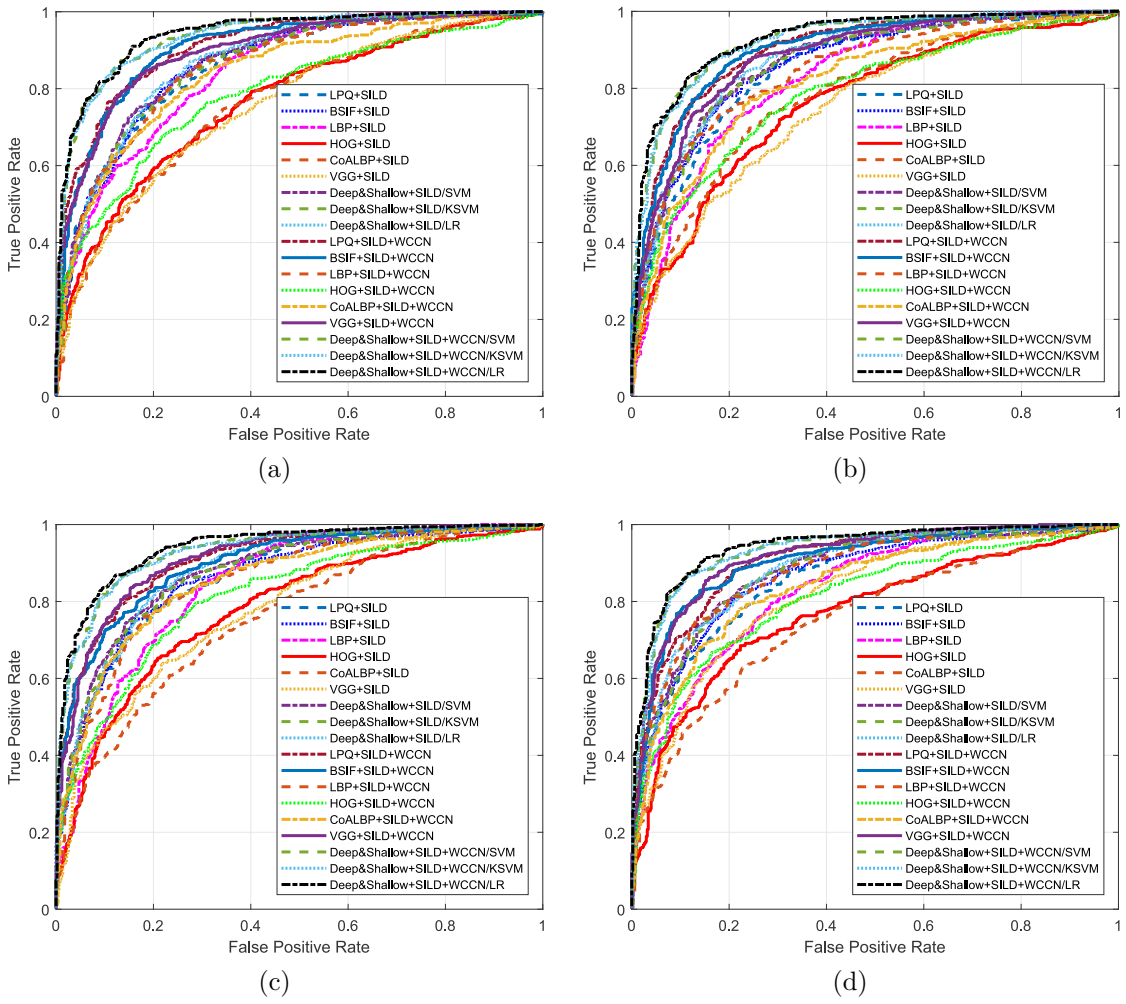


Figure 4.6: ROC curves of different methods (SILD and SILD+WCCN) using the different deep and shallow features on TSKinFace database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

the LFW database (face verification) using MSBSIF description, our SILD+WCCN method performs better than the SILD method with about 8%. Further, for bi-subject kinship verification on both KinFaceW-II and TSKinFace databases using Deep&Shallow description, our SILD+WCCN method outperforms the SILD method with about 5.5% and 6.3% on KinFaceW-II and TSKinFace databases, respectively. For tri-subject kinship verification, our SILD+WCCN method outperforms the SILD method using Deep&Shallow description on TSKinFace database with about 5.56% and 6.18% for FM-S and FM-D relations, respectively. For a general observation, from individual features description and their fusion (i.e. LPQ, BSIF, LBP, HOG, CoALBP, VGG and Deep&Shallow), our SILD+WCCN work well and projects each of these features into new discriminative subspace where their classification becomes easy compared with SILD counterpart. Which means, the integrate of WCCN on SILD play an important role in the enhancement of classification performances through decreasing the within-class variations effect by reducing the expected classification error on the training step.

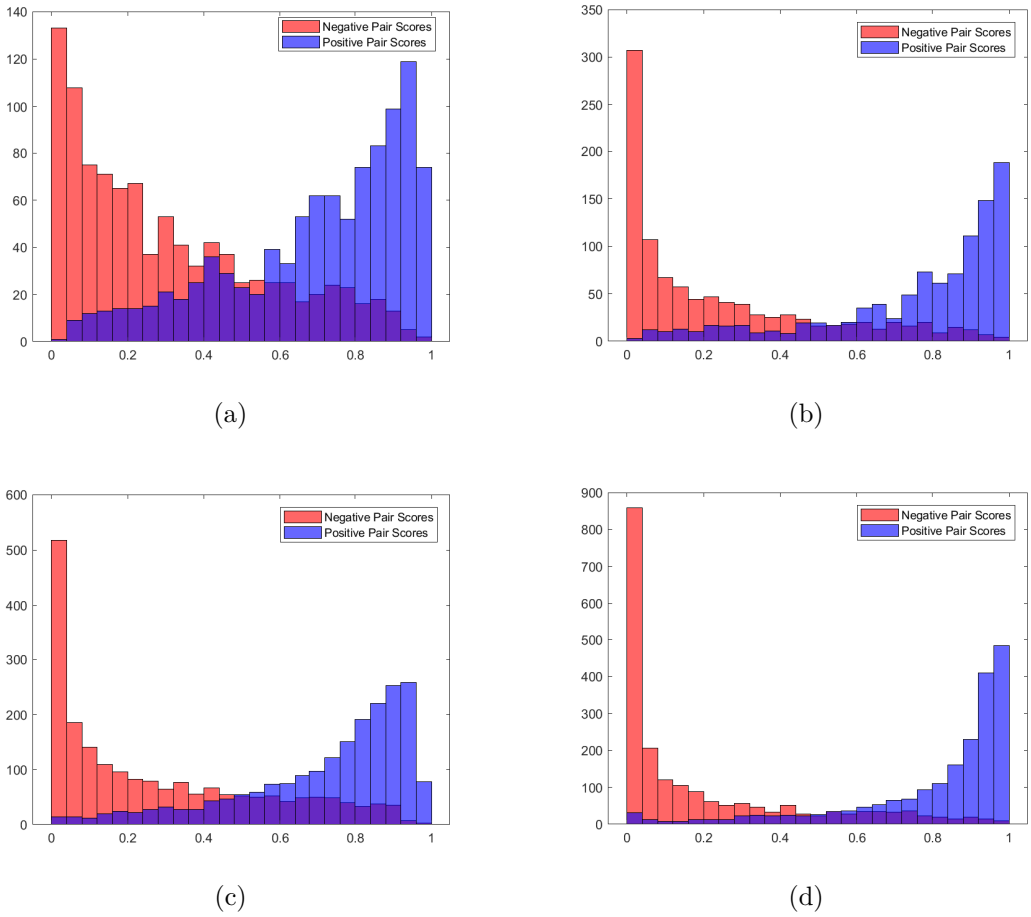


Figure 4.7: The distribution of similarities between the positive pairs (blue) and negative pairs (red) using Deep&Shallow features with, (a) SILD and (b) SILD+WCCN methods on KinFaceW-II database, (c) SILD and (d) SILD+WCCN methods on TSKinFace database.

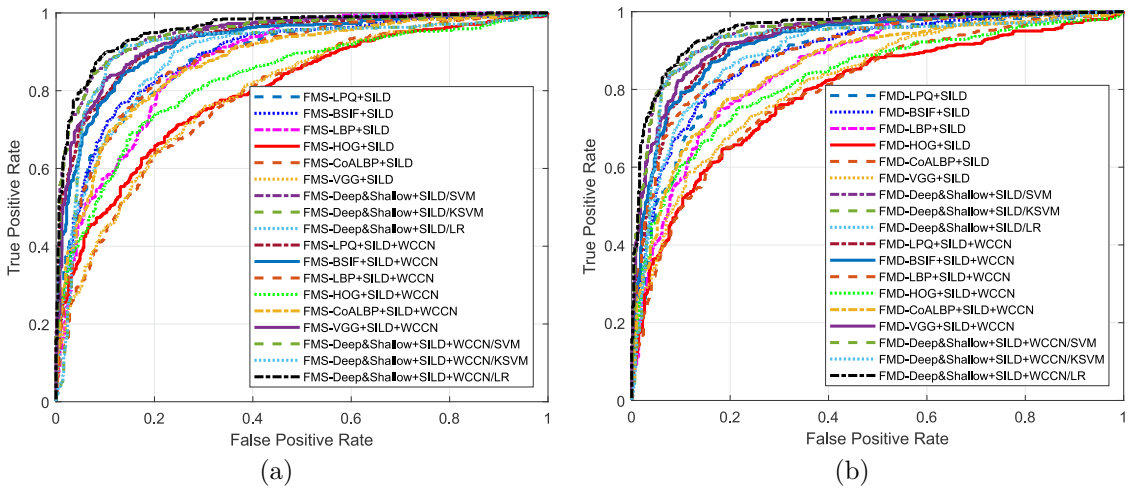


Figure 4.8: ROC curves of different methods (SILD and SILD+WCCN) using the different deep and shallow features on TSKinFace database obtained on (a) FMS relation, (b) FMD relation.

Table 4.4: Performance comparisons (%) with state-of-the-art methods on KinFaceW-II database.

Method	F-S	F-D	M-S	M-D	Mean
MNRML [103]	76.90	74.30	77.40	77.60	76.50
DMML [174]	78.50	76.50	78.50	79.50	78.25
MPDFL [175]	77.30	74.70	77.80	78.00	77.00
DDMML [101]	87.40	83.80	83.20	83.00	84.30
NRCML [172]	79.80	76.10	79.80	80.00	78.70
MKSM [193]	83.80	81.20	82.40	82.40	82.45
MvDML [60]	80.40	79.80	78.80	81.80	80.20
SILD/LR (Deep&Shallow)	85.20	77.40	81.20	79.00	80.70
SILD+WCCN/LR (Deep&Shallow)	**88.40**	**84.20**	**85.80**	**86.40**	**86.20**

Table 4.5: Performance comparisons (%) with state-of-the-art methods on TSKinFace database.

Method	F-S	F-D	M-S	M-D	Mean	FM-S	FM-D
RSBM [133]	83.00	80.50	82.80	81.10	81.85	86.40	84.40
BSIF-HSV [168]	81.47	81.40	79.90	82.00	81.19	/	/
DDMML [101]	86.60	82.50	83.20	84.30	84.15	88.50	87.10
MKSM [193]	84.80	83.20	85.19	84.90	84.52	/	/
SILD/LR (Deep&Shallow)	81.66	81.96	82.65	82.87	82.29	85.38	85.05
SILD+WCCN/LR (Deep&Shallow)	**89.08**	**87.05**	**88.59**	**89.63**	**88.59**	**90.94**	**91.23**

4.4 Conclusion

In this chapter, we presented an effective approach based on deep and shallow features to the problem of kinship verification. To achieve a low dimensional and discriminative subspace, we proposed the SILD+WCCN method. SILD+WCCN finds the projection subspace, where the separation between data classes is enhanced. The experimental evaluation shows the superiority of our method compared with its original form (i.e. SILD method). The best results of our approach are obtained by fusing scores of six multi-view deep and shallow features (Deep&Shallow) projected with the proposed SILD+WCCN/LR method. These results outperform the state of the art on KinFaceW-II and TSKinFace databases for bi-subject matching problem, and outperform the state of the art on the available TSKinFace database for tri-subject matching problem. Furthermore, these results point out to the need of using deep/shallow features for kinship verification. As future work, we plan to investigate the complementarity of more features description for face representation with the proposed dimensionality reduction method (SILD+WCCN/LR).

5 A Weighted Exponential Discriminant Analysis through Side-Information for Face and Kinship Verification using Statistical Binarized Image Features

Contents

5.1 Introduction

Biometric facial images include large human traits, such as identity, gender, expression, age, and ethnicity [11, 106, 123]. Over the past two decades, automatic face recognition under controlled conditions showed increasing results and best performances in wide-scale biometric topics. However, satisfactory performances are still beyond reach because of the big challenge of weakly labeled data setup and under unconstrained environments, in which the latter is characterized by blurring images, low quality, without any restrictions in terms of pose, background, expression, lighting, and partial occlusion. More recently, many public datasets [44, 67, 133, 162, 170] studied face and kinship applications that deal with facial images in unconstrained environments. Most of the applications deal with a scenario of verifying a pair of facial images to check whether the images belong to the same person or to different persons. Therefrom, check the same person who appears in the passport photo is a typical example where identity verification via face can operate, especially by the security organizations [15, 55, 107].

Therefore, pairs of facial images are used for the training stage, over weakly labeled data that denotes if the pair of images belong to the same person or the two facial images belong to two different persons. Whereas in test stage, a new pair of facial images is presented to check the appropriate decision if it is matched/mismatched to the person [66]. Other face application under unconstrained environments through weakly labeled data is checking if two facial images belong to persons from the same family or not. The challenge is to understand and extract the face similarities between family members, because many research works have investigated and proved that facial traits are significant cues to verify the kinship [44, 101, 103, 133, 175].

To tackle this problem, several algorithms have been proposed in the previous literature works to increase the performance of the verification systems. Subspace transformation methods are mostly used for face recognition, which is considered as one of the efficient applied methods in this topic, the mutual role of using these methods is predominantly to transform the high dimensional features into a new lower and discriminative subspace. Which means, enhancing the separation between the classes [190]. The most used and well-known subspace transformation methods are Principle Component Analysis (PCA) [22] and Linear Discriminant Analysis (LDA) [127].

Unsupervised subspace transformation methods cannot properly conserve the features structure model of different classes (i.e the structure of each data class is not taken into consideration). Discriminative features are often conserved by supervised subspace transformation methods. LDA (also called Fisher's linear discriminant or FLD) is the traditional method which learns extracted features and transforms them into discriminant subspace. Unfortunately, it cannot be directly applied because of the small size sample (SSS) problem [8, 136] at the reason of the within-class scatter matrix's singularity. As we

know, face recognition is a typical small size problem. Many works have been reported to use LDA for face recognition. The mostly used approach, called Fisherface (or PCA+LDA), was proposed by Belhumeur *et al.* [12] and Swets *et al.* [145]. In their approach, PCA was first employed to minimize the size of the original feature space from M to d, and then the classical FLD is applied to minimize the dimension from d to n ($n \leq d$). Whereas some useful discriminative information may be lost, caused by the features that are thrown away in the PCA phase. On the other hand, the PCA phase cannot guarantee the success of transforming the within-class scatter matrix to be nonsingular.

Later on, the Exponential Discriminant Analysis (EDA) [190] method is proposed to solve the singularity problem, which has showed interesting results in face recognition, by projecting the features through nonlinear subspace. However, the supervised subspace dimensionality reduction and transformation methods like LDA and EDA need the full labeled class information. Sometimes this information is not available for all classes under the unconstrained settings (i.e. the only information is that pair is match or mismatch/kin or nonkin). Therefore, Kan *et al.* [109], proposed subspace transformation method called Side-information based Linear Discriminant Analysis (SILD), in which they proposed a substitutional solution to compute the between-class scatter matrix (S_b) and within-class scatter matrix (S_w). Besides, defining a new representation that directly can be computed through the side-information (i.e. weakly supervised information, where only the pairwise label information is available to train the methods). Inevitably, SILD method suffers from the SSS problem, thus the authors suggested the solution of Fisherface approach applied in LDA-based methods. Recently, Ouamane *et al.* [121], proposed an effective method named Side-information based Exponential Discriminant Analysis (SIEDA). Inspired by EDA, the authors benefit from the advantages of the matrix exponential which preserved the discriminative information included in the null space of the within-class scatter matrix unlike to PCA+SILD approach. In addition, a kernel method (diffusion) is used for the transformation of the nonlinear posed problems across linear problems. Similarly, to the kernel methods, the within-class and between-class scatter matrices are transformed through a new space by the exponential function [99, 125, 158, 166, 185, 186]. Motivated by this research, we propose in this work a novel Side-information based Weighted Exponential Discriminant Analysis (SIWEDA) method for face and kinship verification in the wild. We reformulate the Fisher criterion of the classical SIEDA to maximize the distance of the between-class scatter matrix and minimize the distance of the within-class scatter matrix using a weighting factor (α) in order to make best separation between classes. Moreover, in unconstrained environment, the applicability of the texture information may be limited by the image degradations. For this reason, feature extraction is a key step that promoted and incubated from the facial recognition research community to describe facial images (i.e. Local Binary Patterns (LBP) [119], Local Phase Quantization (LPQ) [120], Binarized Statistical Image Features (BSIF) [77], Context-Aware Local Binary Feature Learning

(CA-LBFL) [42], Rotation-Invariant Local Binary Descriptor (RI-LBD) [41] and Deep Binary Descriptor with Multi-Quantization (DBD-MQ) [43]). We propose new local face descriptor called Statistical Binarized Image Features (StatBIF). Our descriptor shows more efficient traits unlike the well-known local descriptors in the literature works.

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

1. We introduce a novel local feature for describing facial images. Our descriptor is based on the local statistics traits of the facial image and the original BSIF operator.

2. We propose a new method SIWEDA for face and kinship verification based on the classical SIEDA method. Furthermore, to lighten the class intra-variability, we proposed two variants SIEDA+WCCN and SIWEDA+WCCN by integrating WCCN in SIEDA and SIWEDA, respectively.

3. We extensively evaluate our approach against the state-of-the-art methods using five challenging face and kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, YTF and LFW databases.

The rest of the chapter is organized as follows: Section 5.2 gives demonstration of the proposed descriptor for face and kinship verification. We present and describe our SIWEDA method in Section 5.3. The experimental data and setup are presented and results are discussed in Section 5.4. Finally, concluding remarks are given in Section 5.5.

5.2 Statistical Binarized Image Features (StatBIF)

We propose a new operator to describe kin relation facial images inspired by the statistical operator and the Binarized Statistical Image Features (BSIF) operator [77]. The BSIF operator 5.1 computes an image patch code I of size $l \times l$ pixels and a linear filter Z_i which contains N filters of the same size, the filter response s_i is obtained by

$$s_i = \sum_{m,n} Z_i(m,n) I(m,n) \tag{5.1}$$

In order to enhance the discrimination of BSIF feature to local changes in the facial image and to increase its robustness, we determine a new image space described by the local statistical features extraction. Considering the same method as in LBP operator, each pixel of a given facial image is represented by a statistical value calculated using its neighborhood pixels to form a new image. Thus defined by the number of pixels P

and the circle of radius R. When the new statistical images are produced, the basic BSIF operator, with filter size $l \times l$, is applied to each image separately. The binarized feature b_i is obtained by setting $b_i = 1$ if $s_i > 0$ and $b_i = 0$ otherwise. Formally, the Statistical BIF feature can be obtained by decimal conversion of the binarization of the filters response:

$$StatBIF = \sum_{i=1}^{N} b_i(\sum_{m,n} Z_i(m,n)Stat_{P,R}(m,n)) \times 2^{i-1} \tag{5.2}$$

where P and R are defined as the circle for calculating the statistical value of the pixel where ($R = \frac{l-1}{2}$ and $P = R \times N$) and $l \times l$ is the filter size for computing the BSIF code (see Fig. 5.1). In Eq. 5.2, Stat refers to the statistical function used to generate the new local statistical representation. In our experiments, we consider six statistical functions:

- The median:

$$median_{P,R}(i) = median\left\{i_{p=0}, ..., i_{p=P-1}\right\} \tag{5.3}$$

- The mean:

$$mean_{P,R}(i) = \frac{1}{P}\sum_{p=0}^{P-1} i_p \tag{5.4}$$

- The standard deviation:

$$std_{P,R}(i) = \sqrt{\frac{\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2}{P-1}} \tag{5.5}$$

- The variance:

$$variance_{P,R}(i) = \frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2 \tag{5.6}$$

- The skewness:

$$skewness_{P,R}(i) = \frac{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^3}{\left(\sqrt{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2}\right)^{3/2}} \tag{5.7}$$

- The kurtosis:

$$kurtosis_{P,R}(i) = \frac{\frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^4}{\left(\frac{1}{P}\sum_{p=0}^{P-1}(i_p - mean_{P,R}(i))^2\right)^2} \tag{5.8}$$

The computation of StatBIF features is illustrated in Fig. 5.1. Examples of images obtained by applying StatBIF feature for different statistical functions, on facial image are shown in Fig. 5.2.

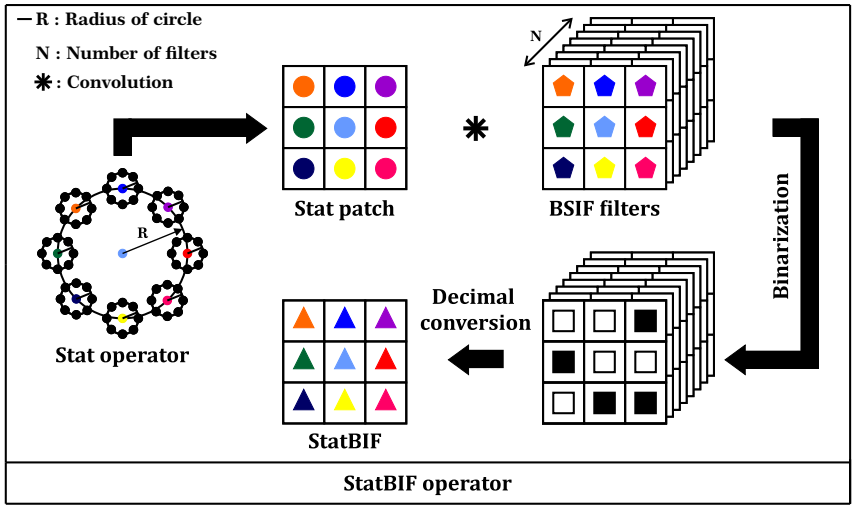


Figure 5.1: Computation of the statistical binarized image features. First, the local statistics of the image are estimated on the circle (P, R). Then, the original BSIF operator, with parameters $l \times l$, is applied.

5.3 Side-Information based Weighted Exponential Discriminant Analysis (SIWEDA)

5.3.1 Matrix Exponential

To introduce SIWEDA method we are going to give the following definitions and properties. Let X be a $q \times q$ square matrix with its exponential denoted by $exp(X)$ or e^X, is defined as:

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

1. $exp(0) = I$.
2. $exp(X)$ is a full rank matrix.
3. If $XY = YX$, then $exp(X+Y) = exp(X)exp(Y) = exp(Y)exp(X)$.
4. $(exp(X))^{-1} = exp(-X)$.

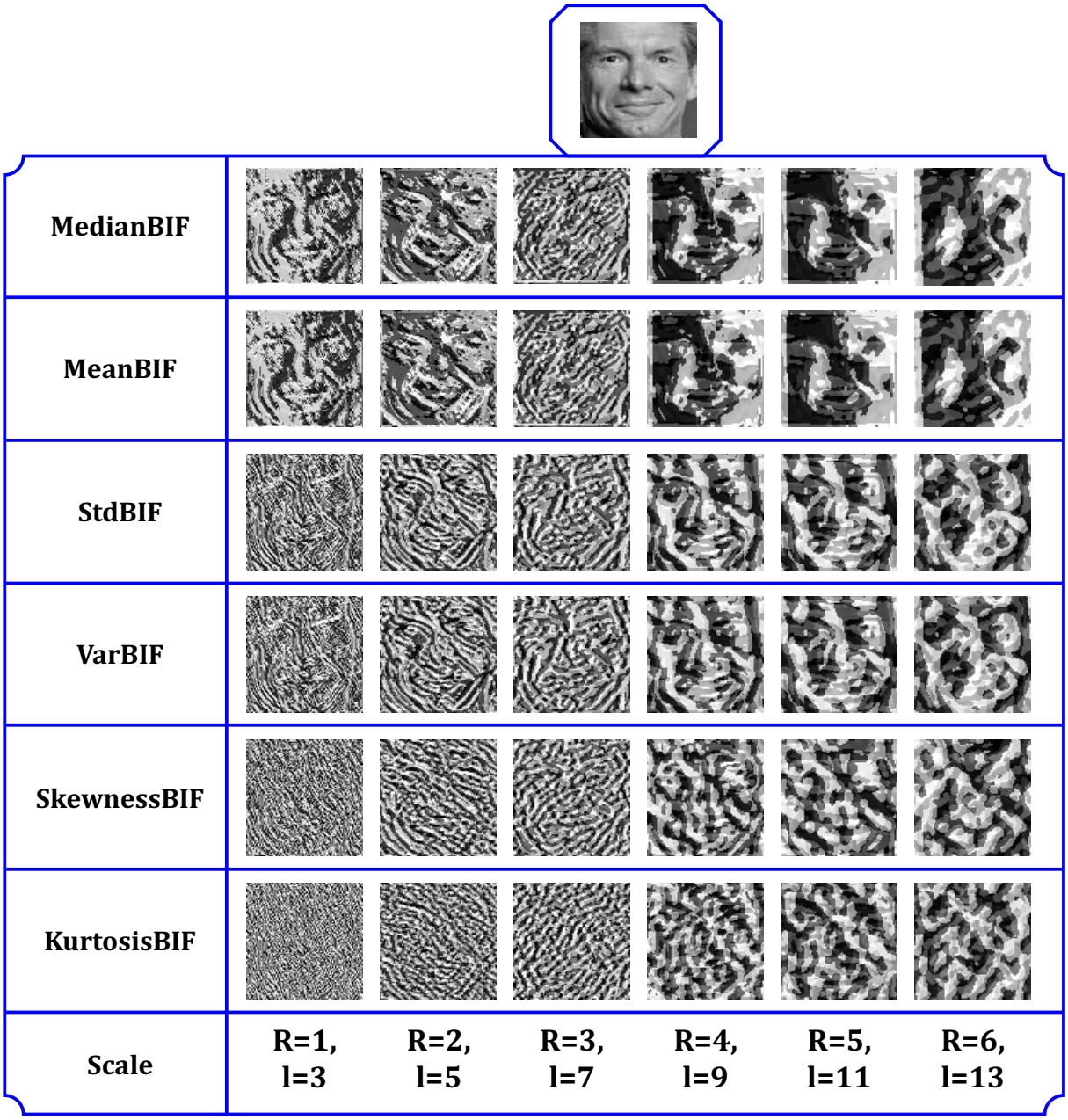


Figure 5.2: Examples of the generated faces by applying the statistical binarized image features, with various scales.

5. For the invertible matrix Z, $exp(Z^{-1}XZ) = Z^{-1}exp(X)Z$.

6. If $X = diag(x_1, x_2, ..., x_m)$ is a diagonal matrix, $exp(X) = diag(exp(x_1), exp(x_2), ..., exp(x_m))$.

7. $|exp(X)| = exp(tr(X))$.

8. If $(v_1, v_2, ..., v_m)$ are eigenvectors for matrix X and $\Lambda_1, \Lambda_2, ..., \Lambda_m$ the corresponding eigenvalues, then $(v_1, v_2, ..., v_m)$ are also eigenvectors of exp(X) that correspond to the eigenvalues $exp(\Lambda_1), exp(\Lambda_2), ..., exp(\Lambda_m)$.

5.3.2 Side-Information based Exponential Discriminant Analysis (SIEDA)

Supervised subspace transformation methods, such as linear discriminant analysis (LDA) [127] and exponential discriminant analysis (EDA) [190], enhance the discrimination of the extracted features by transforming these data in a sub-space where it is easier to perform its separation and classification. These methods need the class information for each sample. The within class scatter matrix (S_w) and the between class scatter matrix (S_b) must be calculated with full label information:

$$S_w = \sum_{i=1}^{L} \frac{1}{n_i} \sum_{j=1}^{n_i} (\xi_j^i - m^i)(\xi_j^i - m^i)^T \tag{5.9}$$

$$S_b = \sum_{i=1}^{L} (m^i - \bar{m})(m^i - \bar{m})^T \tag{5.10}$$

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

That means, LDA and EDA fail in weakly labeled data. Kan et al. [109] proposed a new representation to resolve this problem by directly operating the S_w and S_b matrices with the side-information. The positive classes pair images are directly utilized to calculate the within class scatter matrix and the negative classes pair images are used to compute the between class scatter matrix. Let us refer $P_{class} = \left\{(\check{\xi}_i^1, \hat{\xi}_i^1) : l(\check{\xi}_i^1) = l(\hat{\xi}_i^1)\right\}$ as the collection of positive-class image pairs and $N_{class} = \left\{(\check{\xi}_i^0, \hat{\xi}_i^0) : l(\check{\xi}_i^0) \neq l(\hat{\xi}_i^0)\right\}$ as the collection of negative-class image pairs, where the image ξ is represented by the class label $l(\xi)$. Here, the within-class and between-class scatter matrices of Side-Information based Linear Discriminant analysis (SILD) method can be represented by:

$$S_w^{sild} = \sum_{i=1}^{C_1} (\check{\xi}_i^1 - \hat{\xi}_i^1)(\check{\xi}_i^1 - \hat{\xi}_i^1)^T \tag{5.11}$$

$$S_b^{sild} = \sum_{i=1}^{C_0} (\check{\xi}_i^0 - \hat{\xi}_i^0)(\check{\xi}_i^0 - \hat{\xi}_i^0)^T \tag{5.12}$$

The target function for SILD is:

$$\begin{aligned} U_{opt}^{sild} &= argmax_U \frac{U^T S_b^{sild} U}{U^T S_w^{sild} U} \\ &= argmax_U \frac{\left|U^T (V_b^T \Lambda_b V_b) U\right|}{|U^T (V_w^T \Lambda_w V_w) U|} \end{aligned} \tag{5.13}$$

where:

$$\begin{aligned} S_w^{sild} &= V_w^T \Lambda_w V_w \\ S_b^{sild} &= V_b^T \Lambda_b V_b \end{aligned} \tag{5.14}$$

To make better separation of the discriminative information, Ouamane *et al.* [121] proposed SIEDA method. Like EDA [190], SIEDA replaces the eigenvalues Λ_{w_k} by $exp(\Lambda_{w_k})$ in S_w^{sild}, and the eigenvalues Λ_{b_k} by $exp(\Lambda_{b_k})$ in S_b^{sild}. Thus, from property (8) of the matrix exponential the target function for SILD becomes:

$$\begin{aligned} U_{opt}^{sieda} &= argmax_U \frac{\left|U^T(V_b^T exp(\Lambda_b)V_b)U\right|}{|U^T(V_w^T exp(\Lambda_w)V_w)U|} \\ &= argmax_U \frac{U^T exp(S_b^{sild})U}{U^T exp(S_w^{sild})U} \end{aligned} \tag{5.15}$$

This problem is reduced to a generalized eigenvalue problem:

$$(exp(S_w^{sild}))^{-1} exp(S_b^{sild}) = V^T \mathbf{\Lambda} V \tag{5.16}$$

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

In Eq. (5.13), the inter-class distance of training samples from different subjects $(\check{\xi}_i^0, \hat{\xi}_i^0)|_{i=0}^{C_0}$ for all pairs are maximized by the numerator while the intra-class distance of training samples from the same subjects $(\check{\xi}_i^1, \hat{\xi}_i^1)|_{i=0}^{C_1}$ for all pairs are minimized by the denominator. This equation derived from the following multiobjective programming problem:

$$\begin{aligned} &max \frac{U^T S_b^{sild} U}{U^T U} \\ &min \frac{U^T S_w^{sild} U}{U^T U} \end{aligned} \tag{5.17}$$

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

Whereas, from property (6) of the matrix exponential, the two distances becomes: $dist_b = trace(exp(S_b^{sild})) = exp(\Lambda_{b_1}) + exp(\Lambda_{b_2}) + \cdots + exp(\Lambda_{b_n})$ and $dist_w = trace(exp(S_w^{sild})) = exp(\Lambda_{w_1}) + exp(\Lambda_{w_2}) + \cdots + exp(\Lambda_{w_n})$.

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

5.3.3 Proposed Side-Information based Weighted Exponential Discriminant Analysis (SIWEDA)

To enhance the separation between the positive and negative classes, we propose SIWEDA method to maximize the objective function of SIEDA method. ($\alpha = 1$) is the particular weighting factor value of the objective target function (5.15) of the classical SIEDA. We

generalize the objective target function to contain and accept different values of weighting factor (α).

From (5.13) the generalized objective target function of SILD is:

$$\begin{aligned} U_{opt}^{siwld} &= argmax_U \frac{U^T \alpha S_b^{sild} U}{U^T \alpha S_w^{sild} U} \\ &= argmax_U \frac{\left|U^T \alpha (V_b^T \Lambda_b V_b) U\right|}{|U^T \alpha (V_w^T \Lambda_w V_w) U|} \\ &= argmax_U \frac{\left|U^T (V_b^T \alpha \Lambda_b V_b) U\right|}{|U^T (V_w^T \alpha \Lambda_w V_w) U|} \end{aligned} \tag{5.18}$$

where $\alpha > 0$.

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

$$\begin{aligned} U_{opt}^{siweda} &= argmax_U \frac{\left|U^T (V_b^T exp(\alpha \Lambda_b) V_b) U\right|}{|U^T (V_w^T exp(\alpha \Lambda_w) V_w) U|} \\ &= argmax_U \frac{U^T exp(\alpha S_b^{sild}) U}{U^T exp(\alpha S_w^{sild}) U} \end{aligned} \tag{5.19}$$

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

By applying the weighting factor (α), the distance becomes: $dist_b = trace(exp(\alpha S_b^{sild})) = exp(\alpha \Lambda_{b_1}) + exp(\alpha \Lambda_{b_2}) + \cdots + exp(\alpha \Lambda_{b_n})$ and $dist_w = trace(exp(\alpha S_w^{sild})) = exp(\alpha \Lambda_{w_1}) + exp(\alpha \Lambda_{w_2}) + \cdots + exp(\alpha \Lambda_{w_n})$.

Therefore, the ratio:

$$exp(\alpha \Lambda_{b_k})/exp(\alpha \Lambda_{w_k}) > exp(\Lambda_{b_k})/exp(\Lambda_{w_k}) > \Lambda_{b_k}/\Lambda_{w_k} \tag{5.20}$$

where $\alpha > 1$.

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

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

For more explanation of our method, Figure 5.3 illustrates an example of the proportion of Λ_k, $exp(\Lambda_k)$ and two weighting exponentials $exp(2 \times \Lambda_k)$ and $exp(5 \times \Lambda_k)$ and their respective sums. In this figure, the largest eigenvalue $\Lambda_5 = 33.33\%$ while its corresponding exponential $exp(\Lambda_5) = 63.64\%$, $exp(2 \times \Lambda_5) = 88.47\%$ and $exp(5 \times \Lambda_5) = 99.33\%$ and the smallest eigenvalue $\Lambda_1 = 6.67\%$ while its corresponding exponential $exp(\Lambda_1) = 1.17\%$, $exp(2 \times \Lambda_1) = 0.029\%$ and $exp(5 \times \Lambda_1) = 10^{-7}\%$. We have: $\Lambda_{b_k}/\Lambda_{w_k} < exp(\Lambda_{b_k})/exp(\Lambda_{w_k}) < exp(2 \times \Lambda_{b_k})/exp(2 \times \Lambda_{w_k}) < exp(5 \times \Lambda_{b_k})/exp(5 \times \Lambda_{w_k})$. We see that ($\alpha = 5$) neglected

the Λ_4 which may contain a significant information with sum of 0.67%, unlike to ($\alpha = 2$) which conserve the Λ_4 with sum of 11.70%. This means that, each set of training data from different datasets is adapted with a specific weighting factor (α), that can be simplified the separation between the positive and negative classes.

In section 5.4, our experiments show that by changing the value of the weighting factor (α) gives us the ability for easier selection the eigenvalues of high significance and eliminate those with less discriminative information. Thus, the small eigenvalues are reduced and large eigenvalues are enlarged.

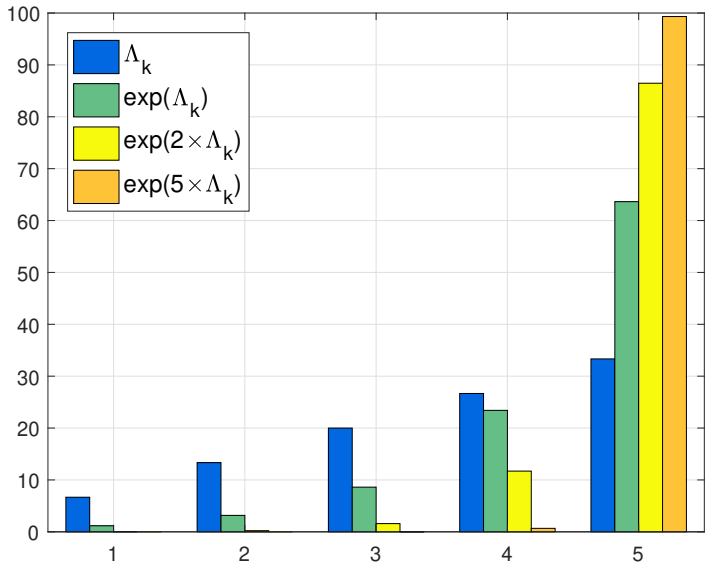


Figure 5.3: Example of the proportions $\frac{\Lambda_k}{\sum \Lambda_k}$ (blue bars), $\frac{exp(\Lambda_k)}{\sum exp(\Lambda_k)}$ (green bars), $\frac{exp(2\times\Lambda_k)}{\sum exp(2\times\Lambda_k)}$ (yellow bars) and $\frac{exp(5\times\Lambda_k)}{\sum exp(5\times\Lambda_k)}$ (orange bars).

5.3.4 Within-class covariance normalization

The first use of the within-class covariance normalization (WCCN) is in the community of speaker recognition. While Dehak *et al.* [34] founded that it is the best technique to project the reduced-vectors of LDA method to a new subspace determined by the square-root of the inverse of the within-class covariance matrix. We propose a new variant of SIWEDA by integrating WCCN:

$$W = \sum_{i=1}^{C_1} \frac{(U^{siweda})^T \check{\xi}_i^1 - (U^{siweda})^T \hat{\xi}_i^1}{(U^{siweda})^T \check{\xi}_i^1 - (U^{siweda})^T \hat{\xi}_i^1} \tag{5.21}$$

where, U^{siweda} is the SIWEDA projection matrix found in Eq.5.19. The WCCN projection matrix C is obtained by Cholesky decomposition of the inverse of W: $W^{-1} = CC^T$. Where the new projection matrix Z^{siweda} is obtained by: $Z^{siweda} = C^T U^{siweda}$. By imposing

upper bounds on the classification error metric [10], WCCN decreases the within-class variations effect by reducing the expected classification error on the training step.

The procedure of this proposed variant, Side-Information Weighted Exponential Discriminant Analysis integrating within class covariance normalization (SIWEDA+WCCN), is detailed in algorithm 2.

Algorithm 2 Side-Information Weighted Exponential Discriminant Analysis plus Within Class Covariance Normalization

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

Output:
- The projection matrix Z^{siweda} of SIWEDA.

Algorithm:

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

5.3.5 Similarity measure

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

$$Cos(\check{\xi}, \hat{\xi}) = \frac{(C^T U^{siweda}\check{\xi})^T (C^T U^{siweda}\hat{\xi})}{\left\|(C^T U^{siweda}\check{\xi})\right\| \left\|(C^T U^{siweda}\hat{\xi})\right\|} \tag{5.22}$$

After discriminant analysis methods, the using of cosine similarity distance has an advantage comes from its connection to the Bayes decision rule, as the optimal used method is the Bayes classifier for decreasing the classification error [95].

5.4 Experiments

The experiments are organized into three parts: First part presents the benchmark datasets utilized in our experiments; Second part gives the parameter settings utilized in

our framework; Third part provides the results with their discuss and compare the best ones with those of the state of the art.

5.4.1 Parameter Settings

Face matching

The LFW dataset is divided into two views: view 1 is utilized for model selection, and view 2 is put to evaluate performance, containing three evaluation paradigms: 1) *the image unrestricted protocol*, 2) *the image restricted protocol* and 3) *the unsupervised protocol*. In our experiments, the aligned images (LFW-a) were used, and the proposed approach was evaluated on the view 2 using image restricted protocol, where no outside additional training data was used. The dataset is subdivided into 10 disjoint folds cross-validation. While for training step, 9 folds are used and the remaining fold for testing step. Each fold includes 300 matched (positive) pairs and 300 mismatched (negative) pairs. The final performance is reported as the mean accuracy ± standard deviation (SE) as well as the ROC curve through the 10-fold cross-validation.

The YouTube Face (YTF) dataset we followed the same evaluation protocol founded in [162] of unconstrained face verification including 5000 video pairs, which were divided into 10 folds and each fold contains 250 positive pairs and 250 negative pairs. We learned the feature representation using MLBP, MLPQ and MBSIF descriptors and our StatBIF descriptor for each frame of video clips, separately. As all facial images have been aligned already, we averaged all descriptors of one video clip to make a mean vector as the feature of the video.

Kinship verification We evaluated the performance of the proposed framework on the same experimental protocol cited in the literature works [103, 174], whither five-fold cross-validation for kinship verification is carried out, while preserving the number of pairs images nearly equal in all folds. By following this protocol, we ensure that our results directly compared to the state of the art. We generated the negative pairs of the kinship randomly such that the appearance of every facial image is only once in the training set. In the training and test phases, the number of negative pairs and positive pairs is equal.

Features extraction Concerning the face normalization, all face attributes descriptions are extracted from images that are aligned and cropped into a resolution of 64×64 pixels. We extract at Multi-scale the Local Binary Patterns (MLBP) [119], the radius R = {1, 2, 3, 4, 5, 6} and the number of pixels in the neighborhood P = 8. For Multi-scale Binarized Statistical Image Features (MBSIF) [77] we use eight filters with different sizes W = {3, 5, 7, 9, 11, 13}. In the Multi-scale Local Phase Quantization (MLPQ) [120], the window size is M = {3, 5, 7, 9, 11, 13}. For Statistical Binarized Image Features (StatBIF), we use six scales, $StatBIF_{l=3}$, $StatBIF_{l=5}$, $StatBIF_{l=7}$, $StatBIF_{l=9}$, $StatBIF_{l=11}$ and $StatBIF_{l=13}$. From each scale the extracted features are the combination of six statistical features of (median, mean, standard deviation, variance, skewness and kurtosis).

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

5.4.2 Results and discussion

To verify our framework, we generate several experiments listed in Tables 5.1 and 5.2. The mean verification accuracies of our SIEDA+WCCN method compared to SIEDA using the proposed StatBIF descriptor and the three existing descriptors MLBP, MLPQ and MBSIF on the experimental datasets is shown in Table 5.1. As can be seen in this table, the best performance for kinship verification using StatBIF descriptor through SIEDA+WCCN method outperforms the use of this descriptor through SIEDA method only. Whereas, our StatBIF descriptor shows the best accuracy with scales, $l = 13$ for Cornell KinFace dataset and $l = 5$ for both, UB KinFace and TSKinFace datasets. As we can see from LFW dataset, SIEDA+WCCN failed to project the StatBIF features into discriminative subspace, because there is a limitation of separation over the specific weighting factor ($\alpha = 1$).

Table 5.1: Mean verification accuracy (%) of our SIEDA+WCCN method compared to the classical SIEDA method using StatBIF, MLBP, MLPQ and MBSIF descriptors on Cornell KinFace, UB KinFace, TSKinFace and LFW.

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

Table 5.2 demonstrates the mean verification accuracy of our SIWEDA+WCCN method using different descriptors with different weighting factors on the experimental datasets. This table shows that the weighting factors ($\alpha > 1$) have the best performance compared with the SIEDA+WCCN method ($\alpha = 1$). Also, there is specific value of (α) that was adapted to the training set of different datasets, which increase the performance of face and kinship verification through the SIWEDA method. We see that by using SIWEDA+WCCN method over different weighting descriptors ($\text{StatBIF}_{l=11,\alpha=5}$),

($\text{StatBIF}_{l=5,\alpha=20}$), ($\text{MBSIF}_{\alpha=15}$), ($\text{StatBIF}_{l=9,\alpha=15}$), and ($\text{StatBIF}_{l=9,\alpha=15}$) show the best accuracies compared to the classical SIEDA+WCCN method ($\alpha = 1$) for Cornell KinFace dataset, UB KinFace dataset, TSKinFace dataset, LFW dataset and YTF dataset, respectively.

Table 5.2: Mean verification accuracy (%) of our SIWEDA+WCCN method using StatBIF descriptor with different weighted factors on Cornell KinFace, UB KinFace, TSKinFace, LFW and YTF datasets.

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

5.4.3 Weighting factor (α) effect

For each specific case of Table 5.2 (regardless the features, datasets, and face matching problem), the proposed approach SIWEDA+WCCN performs better than its classical SIEDA+WCCN counterpart which obviously demonstrates the effectiveness of the proposed SIWEDA+WCCN method. Moreover, our results demonstrate that SIWEDA+WCCN is able to extract better discriminative features than SIEDA+WCCN.

5.4.4 Effect of score fusion

Toward to answer the question: What is the best weighting scale of our StatBIF descriptor can be used? We check the complementarity of different weighting scales (i.e. $\alpha = 15$ and $\alpha = 20$) of StatBIF descriptor. The Multi-scale StatBIF (MStatBIF) score fusion is performed using logistic regression method [56]. We experimented with three datasets LFW and YTF for face matching problem and TSKinFace for kinship verification problem. Table 5.3 shows the comparison of MStatBIF with different weighting scales of our StatBIF descriptor. As we can see from this table, our MStatBIF shows best and stable performance from the three datasets LFW, YTF and TSKinFace. Furthermore, the Cornell KinFace and UB KinFace datasets are gathered with a mixture of four relations types, Father-Son, Father-Daughter, Mother-Son, and Mother-Daughter pair images, unlike to TSKinFace dataset that subdivided into four subsets (i.e. F-S subset, F-D subset, M-S subset, and M-D subset) which make better scores learning with logistic regression method [83, 85].

Table 5.3: Mean verification accuracy of MStatBIF with different weighting scales of StatBIF descriptor on LFW, YTF and TSKinFace datasets.

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

5.4.5 Computational cost

We calculated the computational time needed for the face and kinship verification of one pair face samples using different weighting scales of our SIWEDA+WCCN method. The experiments were implemented using MATLAB 2018a on a PC with an Intel Core i7 2.00 GHz CPU and 8 GB of RAM. The feature extraction for the MLBP, MLPQ, MBSIF and StatBIF descriptors takes 7.3 ms, 37.6 ms, 27.3 ms and 23.1 ms for each sample, respectively. Furthermore, in addition to its robustness, our StatBIF descriptor got the second rank in term of time cost compared with the three well known descriptors. In training step (offline), the estimation of the projection matrices is performed only once. In

the online step, we evaluate the time cost needed by each method to project and match the test pair which is provided in Table 5.4 in ms. This table shows that the best performing method, SIWEDA, runs faster compared with the SIEDA method (i.e. α=1). Furthermore, we see that the time cost of feature extraction is negligible compared to the projection and matching time.

Table 5.4: Time Cost (TC), in ms, taken by different weighting factors for the projection of one pair of facial images.

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

5.4.6 Comparison with the results of the state of the art

- Matching Face pairs in the Wild Table 5.5 shows the comparison of our results under restricted setting protocol with the state of the art methods on the LFW dataset. The corresponding ROC curves of our framework and the state of the art methods on LFW dataset are depicted in Fig. 5.4. The best achieved verification accuracy of our framework is 96.20%, which is very close to the human performance (i.e. 97.53% on LFW dataset). Furthermore, 80 positive and negative pairs facial images from the 6,000 pairs (almost average of 8 positive and negative pairs for each fold) that are misclassified by our approach compared to human performance. Comparing to the state of the art, our result is the first. The framework MRF-Fusion-CSKDA [7] achieves currently the second rank on the LFW dataset in term of verification accuracy. MRF-Fusion-CSKDA is resulted by the fusion of three descriptors which extracted at multi-scale features, MSLBP, MSBSIF, and MSLPQ, and kernel methods. Differently, we utilized only one descriptor (StatBIF), with an effective approach, SIWEDA+WCCN, achieving first rank. Thus, our framework is less complicated and more efficiently computational than the second ranking method. Furthermore, Table 5.6 shows the comparison of our results under restricted setting protocol with the state of the art methods on the YTF dataset. The corresponding ROC curves of our framework and the state of the art methods on YTF dataset are depicted in Fig. 5.5. Thus, our framework got a competitive performance compared to state of the art methods on YTF dataset under restricted setting protocol.

- Kinship Verification in the wild Table 5.7 compares the proposed approach with the state of the art methods on the Cornell KinFace, UB KinFace and TSKinFace datasets. We notice that the best verification accuracy of our framework achieves 78.40% on Cornell KinFace, 76.32% on UB KinFace and 86.37% on TSKinFace. As can be seen from these

results, our approach outperforms the other state of the art methods on two kinship datasets. In addition, we can see that SIWEDA+WCCN improves with a significant margin the kinship performance (more than 4% improvement on Cornell KinFace and UB KinFace). From TSKinFace dataset, we see that the proposed SIWEDA+WCCN method obtains the first best results for all relationships.

Our approach vs. DeepFace [146] (using 4.4M outside data) We compared our approach using the provided data only on LFW dataset with DeepFace [146] method which used 4.4 Millions outside facial images belonging to more than 4,000 identities for training the network model. We got a competitive performance compared with DeepFace method under restricted setting on LFW dataset (we refer that our StatBIF+SIWEDA+WCCN method and DeepFace method got a verification accuracies of 96.20% and 97.15%, respectively). Moreover, the authors in [146] combine three networks feeding by three types of inputs (i.e. a 3D-aligned 3-channels (RGB) facial images; a gray facial images plus image gradient magnitude and orientation; and a 2D-aligned RGB facial images) of size 152 by 152, unlike to our approach (StatBIF+SIWEDA+WCCN) which used gray input facial images of size 64 by 64 with no preprocessing stage applied to the facial images (a raw facial images was used).

Our approach vs. Deep multi-metric learning [101] In this comparison, we focus on three datasets LFW, YTF and TSKinFace. We see that our approach improves the performance with about 3% from LFW dataset, got a competitive performance from YTF dataset, and improves the performance with about 2% from TSKinFace dataset compared with DDMML. The Discriminative Deep Multi-Metric Learning (DDMML) [101] method used the combination of multiple features (multi-view features) to describe the facial images. The work of Lu et al. [101] adopts the different sizes to extract the features of the input facial images for each dataset (i.e. 80×150 for LFW dataset, 100×100 for YTF dataset, and 64×64 for TSKinFace dataset). Furthermore, they extract different features for each dataset, which are six original and square root features of Sparse SIFT (SSIFT) [53] on the "funneled" LFW dataset, histogram of oriented gradients (HOG) [29] on the LFW-a dataset and high-dimensional LBP (HDLBP) [24] on the original LFW (i.e. they used the combination of six of original and square root features applied on three versions of LFW dataset). For YTF dataset, three features description including LBP [119], Center-Symmetric LBP (CSLBP) [162] and Four-Patch LBP (FPLBP) [163] are used as in [162]. For TSKinFace dataset, they used four descriptors LBP, Dense SIFT (DSIFT), HOG and LPQ. On the other hand, our approach used efficient methods (SIWEDA+WCCN) using the proposed descriptor (StatBIF) only. Furthermore, our approach used the same parameter settings of the all four used descriptors in our work for both face matching and kinship verification problems (i.e. for each dataset the input size is 64×64 and we used the same features extraction parameters for MLBP, MBSIF, MLPQ, and MStatBIF descriptors on each dataset) attains stable, robust and good performances.

Therefore, the DDMML method accepts only a specific input size of facial images and adapted descriptors (under human surveillance/guidance learning) for each dataset, unlike SIWEDA+WCCN do. Besides, this means that the DDMML method has weak learning compared with our SIWEDA+WCCN method for face and kinship verification in all scenarios.

Table 5.5: Comparison verification accuracy of StatBIF+SIWEDA+WCCN with image restricted setting (no outside training data was used) on LFW dataset.

Method	Mean Accuracy ± Standard Error (%)
Eigenfaces, original [150]	60.02 ± 0.79
Nowak2, original [118]	72.45 ± 0.40
Nowak2, funneled [65]	73.93 ± 0.49
Hybrid descriptor-based, funneled [163]	78.47 ± 0.51
3x3 Multi-Region Histograms [140]	72.95 ± 0.55
Pixels/MKL, funneled [130]	68.22 ± 0.41
V1-like/MKL, funneled [130]	79.35 ± 0.55
APEM (fusion), funneled [90]	84.08 ± 1.20
MRF-MLBP [6]	79.08 ± 0.14
Fisher vector faces [142]	87.47 ± 1.49
Eigen-PEP [91]	88.97 ± 1.32
MRF-MBSIF-CSKDA [7]	93.63 ± 1.27
MRF-Fusion-CSKDA [7]	95.89 ± 1.94
POP-PEP [89]	91.10 ± 1.47
DDML [59]	90.68 ± 1.41
LM^{3}L [63]	89.57 ± 1.53
Discriminative deep multi-metric learning [101]	93.28 ± 0.39
CA-LBFL [42]	92.75 ± 1.13
StatBIF-SIWEDA-WCCN (Our)	**96.20 ± 0.63**

Table 5.6: Comparison verification accuracy of StatBIF+SIWEDA+WCCN with image restricted setting (no outside training data was used) on YTF dataset.

Method	Mean Accuracy ± Standard Error (%)
MBGS L2 mean [162]	76.40 ± 1.80
APEM-FUSION [90]	79.10 ± 1.50
STFRD+PMML [27]	79.50 ± 2.50
VSOF+OSS [115]	79.70 ± 1.80
DDML [59]	82.34 ± 1.47
LM^{3}L [63]	81.30 ± 1.20
Discriminative deep multi-metric learning [101]	82.54 ± 1.58
CA-LBFL [42]	**83.30 ± 1.30**
StatBIF-SIWEDA-WCCN (Our)	**80.24 ± 0.47**

Table 5.7: Comparison verification accuracy of StatBIF+SIWEDA+WCCN with state of the art on the Cornell KinFace, UB KinFace and TSKinFace datasets.

Method	Cornell	UB KinFace	TSKinFace
Pictorial structure model [44]	70.67	/	/
Transfer subspace learning [170]	/	68.50	/
Neighborhood repulsed metric learning [103]	71.60	67.05	/
Discriminative multimetric learning [174]	73.50	72.25	/
Prototype discriminative feature learning [175]	71.90	67.30	/
Relative symmetric bilinear model [133]	/	/	81.85
BSIF-HSV [168]	/	/	81.19
Discriminative deep multi-metric learning [101]	/	/	84.15
MHDL3 - {HOG + Color + LPQ} [105]	76.60	/	/
Heterogeneous similarity learning [131]	68.40	56.20	/
StatBIF-SIWEDA-WCCN (Our)	**78.40**	**76.32**	**86.37**

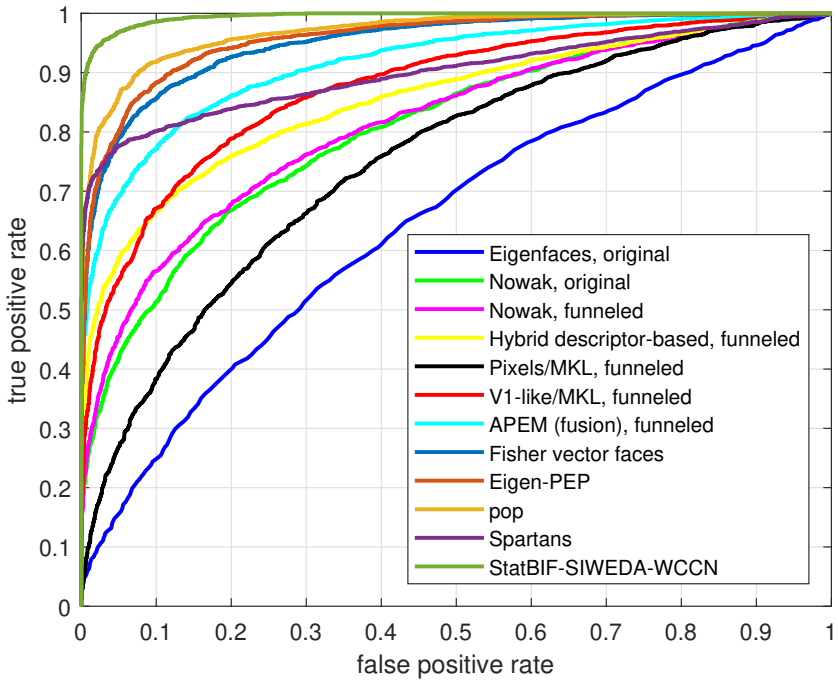


Figure 5.4: ROC curve of StatBIF-SIWEDA-WCCN and other state-of-the-art methods on the LFW dataset under image restricted setting.

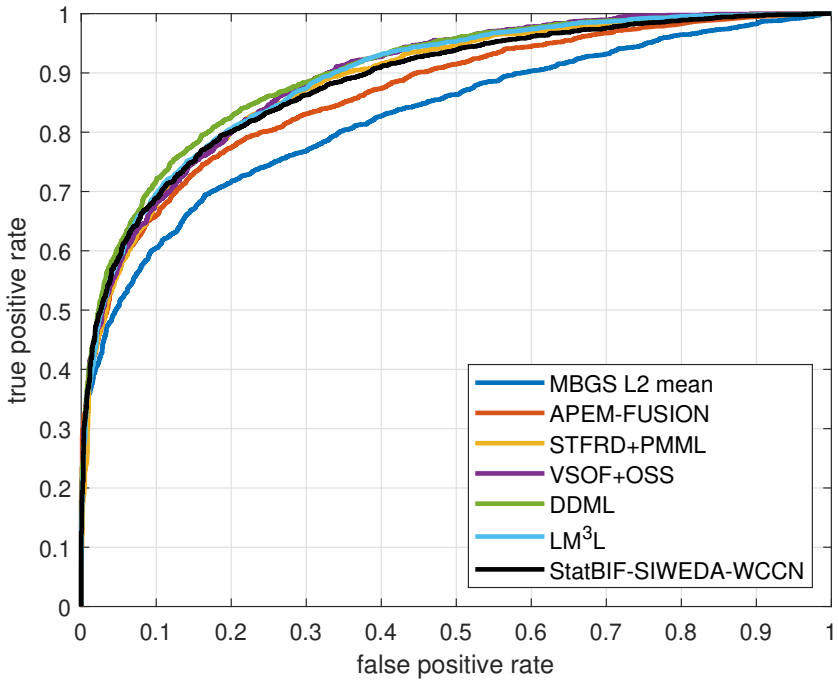


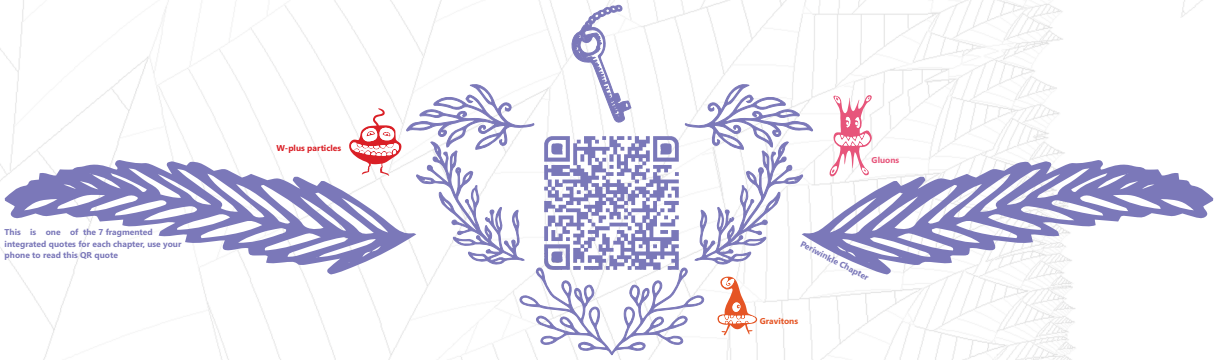
Figure 5.5: ROC curve of StatBIF-SIWEDA-WCCN and other state-of-the-art methods on the YTF dataset under image restricted setting.

5.5 Conclusion

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

6 Tensor Cross-view Quadratic Discriminant Analysis for Kinship Verification in the Wild

Contents



6.1 Introduction

Many researchers [101, 103, 105, 133, 134, 168, 174, 175, 191, 193] used metric learning methods and have achieved reasonably good performance in kinship verification, but none of these methods tackle the kinship verification as a cross-view matching problem.

To our best knowledge, our work is the first effort that tackles the kinship verification problem with a method used in the cross-view matching problem that arise from many applications like heterogeneous face recognition [94] and viewpoint invariant person re-identification [50]. The Cross-view Quadratic Discriminant Analysis (XQDA) [93] method shows the best performances in person re-identification field. Motivated by this research, we propose Tensor Cross-view Quadratic Discriminant Analysis (TXQDA) to analyze the multifactor structure of face images which is related to kinship, age, gender, expression, illumination and pose.

In our framework, the set of face images are represented as a third-order tensor based on local histogram features of the local descriptors, Multi-Scale Local Phase Quantization [120], and Multi-Scale local Binarised Statistical Image Features [77]. The contributions of this work are summarized as follows:

1. We tackle for the first time the kinship verification problem as a cross-view matching problem because every kin relation is typically viewpoint changes from two face images belonging to two different persons.

2. We propose a robust automated facial verification framework suitable for kinship verification, from face images captured in unconstrained environments. The face data is represented as a high order tensor based on the combination of different local features in order to provide a more powerful face model.

3. We propose a novel method for dimensionality reduction and classification, called Tensor Cross-view Quadratic Discriminant Analysis (TXQDA), which preserves the data structure, enlarges the margin between samples, helps lighten the small sample size problem and reduced the computational cost.

4. We extensively evaluate our TXQDA method against the state-of-the-art methods using five challenging kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW.

The chapter is organized as follows: Section 6.2 describes the proposed Tensor Cross-view Quadratic Discriminant Analysis. Our tensor kinship verification pipeline is presented in Section 6.3. The experiments and results are given in Section 6.4. Finally, concluding remarks are given in Section 6.5.

6.2 Tensor Cross-view Quadratic Discriminant Analysis

The variables and mathematical notations that we used in our work are as follows : Lowercase and uppercase symbols (e.g., i, j, F, N and V) indicate scalars; Bold lowercase symbols (e.g., $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$) indicate vectors; italic uppercase symbols (e.g., U, X, Y and W) indicate matrices; bold italic uppercase symbols (e.g., $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$) indicate tensors. A tensor is explained as a multidimensional array [82,179]. N is considered the order of the tensor and $\boldsymbol{X}$ is called an N^{th}-order tensor. $\mathrm{I_k}$, $1 \leq \mathrm{k} \leq \mathrm{N}$, is the dimension of the k^{th} mode. For more details on concepts of tensor algebra, see Appendix C.

6.2.1 Cross-view Quadratic Discriminant Analysis (XQDA)

XQDA [93] is the extended method of the Bayesian face [111] and KISSME [81] approaches to cross-view metric learning, where considered to learn a subspace $W = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_\mathrm{r}) \in \Re^{\mathrm{d}\times\mathrm{r}}$ with cross-view (i.e. Parent-Child) data, and learn a distance function in the r dimensional subspace for the cross-view similarity measure at the same time. We assume that we have a cross-view training set $\{X, Z\}$ of c classes, in which $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathrm{n}) \in \Re^{\mathrm{d}\times\mathrm{n}}$ includes n samples in a d-dimensional space from one view (i.e. Parents samples), $Z = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_\mathrm{m}) \in \Re^{\mathrm{d}\times\mathrm{m}}$ includes m samples in the same d-dimensional space but from the other view (i.e. Children samples). Note that Z is the same with X in the single-view matching scenario. Considering a subspace W, the distance function in the r dimensional subspace is computed as:

$$\mathrm{d}_W(\mathrm{x},\mathrm{z}) = (\mathrm{x}-\mathrm{z})^T W(\Sigma_I^{'-1} - \Sigma_E^{'-1})W^T(\mathrm{x}-\mathrm{z}) \tag{6.1}$$

Where $\Sigma_I^{'} = W^T\Sigma_I W$ and $\Sigma_E^{'} = W^T\Sigma_E W$. Then, we learn a kernel matrix $M(W) = W(\Sigma_I^{'-1} - \Sigma_E^{'-1})W^T$. In [93] the projection direction W is optimized such that $\Sigma_E^{'}/\Sigma_I^{'}$ is maximized. Consequently, $\Sigma_E^{'}/\Sigma_I^{'}$ corresponds to the Generalized Rayleigh Quotient:

$$J(W) = \frac{W^T\Sigma_E W}{W^T\Sigma_I W} \tag{6.2}$$

The two covariance matrices Σ_E and Σ_I are computed as follow:

$$\mathrm{n_I}\Sigma_I = \tilde{X}\tilde{X}^T + \tilde{Z}\tilde{Z}^T - SR^T - RS^T \tag{6.3}$$

Where $\tilde{X} = (\sqrt{\mathrm{m_1}}\mathbf{x}_1, \sqrt{\mathrm{m_1}}\mathbf{x}_2, \ldots, \sqrt{\mathrm{m_1}}\mathbf{x}_{\mathrm{n_1}}, \ldots, \sqrt{\mathrm{m_c}}\mathbf{x}_\mathrm{n})$,
$\tilde{Z} = (\sqrt{\mathrm{n_1}}\mathbf{z}_1, \sqrt{\mathrm{n_1}}\mathbf{z}_2, \ldots, \sqrt{\mathrm{n_1}}\mathbf{z}_{\mathrm{m_1}}, \ldots, \sqrt{\mathrm{n_c}}\mathbf{z}_\mathrm{m})$,
$S = (\sum_{\mathrm{y_i}=1}\mathbf{x}_\mathrm{i}, \sum_{\mathrm{y_i}=2}\mathbf{x}_\mathrm{i}, \ldots, \sum_{\mathrm{y_i}=\mathrm{c}}\mathbf{x}_\mathrm{i})$, $R = (\sum_{\mathrm{l_j}=1}\mathbf{z}_\mathrm{j}, \sum_{\mathrm{l_j}=2}\mathbf{z}_\mathrm{j}, \ldots, \sum_{\mathrm{l_j}=\mathrm{c}}\mathbf{z}_\mathrm{j})$, $\mathrm{y_i}$ and $\mathrm{l_j}$ are class labels, $\mathrm{n_i}$ is the number of samples in class i of X, and $\mathrm{m_i}$ is the number of samples in class i of Z.

$$\mathrm{n_E}\Sigma_E = \mathrm{m}XX^T + \mathrm{n}ZZ^T - \mathbf{s}\mathbf{r}^T - \mathbf{r}\mathbf{s}^T - \mathrm{n_I}\Sigma_I \tag{6.4}$$

Where $\mathbf{s} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathbf{x}_{\mathrm{i}}$ and $\mathbf{r} = \sum_{\mathrm{j}=1}^{\mathrm{m}} \mathbf{z}_{\mathrm{j}}$

6.2.2 Tensor Cross-view Quadratic Discriminant Analysis (TXQDA)

Let a Tensor cross-view training set $\{\boldsymbol{X}, \boldsymbol{Z}\}$ of c classes, where: $\boldsymbol{X} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_{\mathrm{N}} \times \mathrm{n}}$ contains n samples of one view (Parents samples) and $\boldsymbol{Z} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_{\mathrm{N}} \times \mathrm{m}}$ contains m samples of other view (Children samples). The goal of our TXQDA is the calculation of N projection matrices ($W_1 \in \Re^{\mathrm{I}_1 \times \mathrm{I}_1^{'}}, W_2 \in \Re^{\mathrm{I}_2 \times \mathrm{I}_2^{'}}, \ldots, W_{\mathrm{N}} \in \Re^{\mathrm{I}_{\mathrm{N}} \times \mathrm{I}_{\mathrm{N}}^{'}}$). Thus, we calculate one projection matrix for each tensor mode. The objective function of XQDA 6.2 is transformed into:

$$J(W_{\mathrm{k}}) = \frac{W_{\mathrm{k}}^T \Sigma_E^{\mathrm{k}} W_{\mathrm{k}}}{W_{\mathrm{k}}^T \Sigma_I^{\mathrm{k}} W_{\mathrm{k}}} \tag{6.5}$$

We calculate the two covariance matrices Σ_E^{k} and Σ_I^{k} for each k mode by:

$$\mathrm{n_I}\Sigma_I = \sum_{\mathrm{p}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_I}\Sigma_I^{\mathrm{p}}, \mathrm{n_I}\Sigma_I^{\mathrm{p}} = \tilde{X}^{\mathrm{k,p}}(\tilde{X}^{\mathrm{k,p}})^T + \tilde{Z}^{\mathrm{k,p}}(\tilde{Z}^{\mathrm{k,p}})^T - S^{\mathrm{k,p}}(R^{\mathrm{k,p}})^T - R^{\mathrm{k,p}}(S^{\mathrm{k,p}})^T \tag{6.6}$$

Where $\tilde{X}^{\mathrm{k,p}} = (\sqrt{\mathrm{m_1}}\mathbf{x}_1^{\mathrm{k,p}}, \sqrt{\mathrm{m_1}}\mathbf{x}_2^{\mathrm{k,p}}, \ldots, \sqrt{\mathrm{m_1}}\mathbf{x}_{\mathrm{n_1}}^{\mathrm{k,p}}, \ldots, \sqrt{\mathrm{m_c}}\mathbf{x}_{\mathrm{n}}^{\mathrm{k,p}})$,
$\tilde{Z}^{\mathrm{k,p}} = (\sqrt{\mathrm{n_1}}\mathbf{z}_1^{\mathrm{k,p}}, \sqrt{\mathrm{n_1}}\mathbf{z}_2^{\mathrm{k,p}}, \ldots, \sqrt{\mathrm{n_1}}\mathbf{z}_{\mathrm{m_1}}^{\mathrm{k,p}}, \ldots, \sqrt{\mathrm{n_c}}\mathbf{z}_{\mathrm{m}}^{\mathrm{k,p}})$,
$S^{\mathrm{k,p}} = (\sum_{\mathrm{y_i}=1} \mathbf{x}_{\mathrm{i}}^{\mathrm{k,p}}, \sum_{\mathrm{y_i}=2} \mathbf{x}_{\mathrm{i}}^{\mathrm{k,p}}, \ldots, \sum_{\mathrm{y_i}=\mathrm{c}} \mathbf{x}_{\mathrm{i}}^{\mathrm{k,p}})$, $R^{\mathrm{k,p}} = (\sum_{\mathrm{l_j}=1} \mathbf{z}_{\mathrm{j}}^{\mathrm{k,p}}, \sum_{\mathrm{l_j}=2} \mathbf{z}_{\mathrm{j}}^{\mathrm{k,p}}, \ldots, \sum_{\mathrm{l_j}=\mathrm{c}} \mathbf{z}_{\mathrm{j}}^{\mathrm{k,p}})$,
Where, for all presentations, $\mathbf{x}^{\mathrm{k,p}}$ and $\mathbf{z}^{\mathrm{k,p}}$ are the p^{th} column vectors of the k-mode unfolded matrices X^{k} and Z^{k} of sample tensors $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively.

$$\mathrm{n_E}\Sigma_E = \sum_{\mathrm{p}=1}^{\prod_{\mathrm{o}\neq\mathrm{k}} \mathrm{I_o}} \mathrm{n_E}\Sigma_E^{\mathrm{p}}, \mathrm{n_E}\Sigma_E^{\mathrm{p}} = \mathrm{m}X^{\mathrm{k,p}}(X^{\mathrm{k,p}})^T + \mathrm{n}Z^{\mathrm{k,p}}(Z^{\mathrm{k,p}})^T - \mathbf{s}^{\mathrm{k,p}}(\mathbf{r}^{\mathrm{k,p}})^T - \mathbf{r}^{\mathrm{k,p}}(\mathbf{s}^{\mathrm{k,p}})^T - \mathrm{n_I}\Sigma_I^{\mathrm{p}} \tag{6.7}$$

Where $X^{\mathrm{k,p}} = (\mathbf{x}_1^{\mathrm{k,p}}, \mathbf{x}_2^{\mathrm{k,p}}, \ldots, \mathbf{x}_{\mathrm{n_1}}^{\mathrm{k,p}}, \ldots, \mathbf{x}_{\mathrm{n}}^{\mathrm{k,p}})$, $Z^{\mathrm{k,p}} = (\mathbf{z}_1^{\mathrm{k,p}}, \mathbf{z}_2^{\mathrm{k,p}}, \ldots, \mathbf{z}_{\mathrm{m_1}}^{\mathrm{k,p}}, \ldots, \mathbf{z}_{\mathrm{m}}^{\mathrm{k,p}})$, $\mathbf{s}^{\mathrm{k,p}} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathbf{x}_{\mathrm{i}}^{\mathrm{k,p}}$ and $\mathbf{r}^{\mathrm{k,p}} = \sum_{\mathrm{j}=1}^{\mathrm{m}} \mathbf{z}_{\mathrm{j}}^{\mathrm{k,p}}$

Now that the solution for one mode is known, the optimization problem in equation 6.5 can be solved iteratively. The projection matrices $W_1, W_2, \ldots, W_{\mathrm{N}}$ are first initialized to identity. At each iteration $W_1, W_2, \ldots, W_{\mathrm{k}-1}, W_{\mathrm{k}+1}, \ldots W_{\mathrm{N}}$ are hypothetical known and W_{k} is estimated. Set: $\boldsymbol{U} = \boldsymbol{X} \times_1 W_1 \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1} \ldots \times_{\mathrm{N}} W_{\mathrm{N}}$ and $\boldsymbol{Y} = \boldsymbol{Z} \times_1 W_1 \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1} \ldots \times_{\mathrm{N}} W_{\mathrm{N}}$ are replaced in equation 6.5 by $\boldsymbol{X}$ and $\boldsymbol{Z}$. The new equation can be solved by the generalized eigenvalue decomposition problem:

$$\Sigma_E^{\mathrm{k}} W_{\mathrm{k}} = \Lambda_{\mathrm{k}} \Sigma_I^{\mathrm{k}} W_{\mathrm{k}} \tag{6.8}$$

Where, W_{k} is the eigenvectors matrix and Λ_{k} the eigenvalues matrix.

The iterative process of TXQDA breaks up on the recognition of one of the following situations: i) The number of iterations reaches a predefined maximum; or ii) the difference of the estimated projection between two consecutive iterations is less than a threshold, $\left\|W_{\mathrm{k}}^{\mathrm{iter}} - W_{\mathrm{k}}^{\mathrm{iter}-1}\right\| < \mathrm{I_k I_k}\epsilon$, where $\mathrm{I_k}$ is the k mode dimension of $W_{\mathrm{k}}^{\mathrm{iter}}$. The number of iterations, for our TXQDA algorithm, is empirically tuned and the better value is $\mathrm{Iteration_{max}} = 2$.

We summarize the advantages of our algorithm, Tensor cross-view quadratic discriminant analysis (TXQDA), as follows:

1. TXQDA preserves the data structure, where these data stacked in a tensor mode providing the maximum extraction of information. Unlike in the case of XQDA method, the feature vectors are purely concatenated neglecting the natural structure of data.

2. TXQDA also helps lightening the small sample size problem. This is an intrinsic limitation of the XQDA when applying the histograms concatenation of local descriptors for all face blocks (the features length is larger than the number of training samples)

3. TXQDA is a cross-view dimensionality reduction method. It can obviate the curse of dimensionality dilemma by using higher order tensors and k-mode optimization approach, where the latter is performed in a much lower-dimension feature space than the traditional vector-based methods, such as XQDA, do.

4. Many more feature dimensions are available in TXQDA than in XQDA because the available feature dimension of XQDA is theoretically limited by the number of classes in the data, whereas the TXQDA is not.

5. TXQDA reduces the computational cost to a large extent, as the k-mode optimization in each step is performed on a feature space of smaller size.

Consequently, the classification with the proposed TXQDA is better than XQDA. The entire procedure for the proposed Tensor Cross-view Quadratic Discriminant Analysis (TXQDA) is provided in Algorithm 3. The input of this algorithm is defined as follow:

- The tensor $\boldsymbol{X} \in \Re^{\mathrm{I_1}\times\mathrm{I_2}\times\cdots\times\mathrm{I_N}\times\mathrm{n}}$ contains n samples of one view (Parents samples).
- The tensor $\boldsymbol{Z} \in \Re^{\mathrm{I_1}\times\mathrm{I_2}\times\cdots\times\mathrm{I_N}\times\mathrm{m}}$ contains m samples of other view (Children samples).
- $\mathrm{Iteration_{max}}$ is the maximal number of iterations.
- The final lower dimensions: $\acute{\mathrm{I}}_1 \times \acute{\mathrm{I}}_2 \times \cdots \times \acute{\mathrm{I}}_{\mathrm{N}}$.

Whereas the output can defined as follow:

- The projection matrices $W_{\mathrm{k}} = W_{\mathrm{k}}^{\mathrm{iter}} \in \Re^{\mathrm{I_k} \times \acute{\mathrm{I}}_{\mathrm{k}}}, \mathrm{k} = 1, \cdots, \mathrm{N}$

Algorithm 3 Tensor Cross-view Quadratic Discriminant Analysis (TXQDA)

Input:

- $\boldsymbol{X} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_{\mathrm{N}} \times \mathrm{n}}$
- $\boldsymbol{Z} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_{\mathrm{N}} \times \mathrm{m}}$
- $\mathrm{Iteration}_{\max}$
- $\acute{\mathrm{I}}_1 \times \acute{\mathrm{I}}_2 \times \cdots \times \acute{\mathrm{I}}_{\mathrm{N}}$

Output:

- $W_{\mathrm{k}} = W_{\mathrm{k}}^{\mathrm{iter}} \in \Re^{\mathrm{I_k} \times \acute{\mathrm{I}}_{\mathrm{k}}}, \mathrm{k} = 1, \cdots, \mathrm{N}$

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
      - $\boldsymbol{U} = \boldsymbol{X} \times_1 W_1^{\mathrm{iter}-1} \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1}^{\mathrm{iter}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1}^{\mathrm{iter}-1} \ldots \times_{\mathrm{N}} W_{\mathrm{N}}^{\mathrm{iter}-1}$
      - $U^{\mathrm{k}} \Leftarrow_{\mathrm{k}} \boldsymbol{U}$
      - $\boldsymbol{Y} = \boldsymbol{Z} \times_1 W_1^{\mathrm{iter}-1} \ldots \times_{\mathrm{k}-1} W_{\mathrm{k}-1}^{\mathrm{iter}-1} \times_{\mathrm{k}+1} W_{\mathrm{k}+1}^{\mathrm{iter}-1} \ldots \times_{\mathrm{N}} W_{\mathrm{N}}^{\mathrm{iter}-1}$
      - $Y^{\mathrm{k}} \Leftarrow_{\mathrm{k}} \boldsymbol{Y}$
      - $\mathrm{n_I}\Sigma_I = \sum_{\mathrm{p}=1}^{\prod_{\mathrm{o} \neq \mathrm{k}} \mathrm{I_o}} \mathrm{n_I}\Sigma_I^{\mathrm{p}}, \mathrm{n_I}\Sigma_I^{\mathrm{p}} = \tilde{U}^{\mathrm{k,p}}(\tilde{U}^{\mathrm{k,p}})^T + \tilde{Y}^{\mathrm{k,p}}(\tilde{Y}^{\mathrm{k,p}})^T - S^{\mathrm{k,p}}(R^{\mathrm{k,p}})^T - R^{\mathrm{k,p}}(S^{\mathrm{k,p}})^T$
      - $\mathrm{n_E}\Sigma_E = \sum_{\mathrm{p}=1}^{\prod_{\mathrm{o} \neq \mathrm{k}} \mathrm{I_o}} \mathrm{n_E}\Sigma_E^{\mathrm{p}}, \mathrm{n_E}\Sigma_E^{\mathrm{p}} = \mathrm{m}U^{\mathrm{k,p}}(U^{\mathrm{k,p}})^T + \mathrm{n}Y^{\mathrm{k,p}}(Y^{\mathrm{k,p}})^T - \mathbf{s}^{\mathrm{k,p}}(\mathbf{r}^{\mathrm{k,p}})^T - \mathbf{r}^{\mathrm{k,p}}(\mathbf{s}^{\mathrm{k,p}})^T - \mathrm{n_I}\Sigma_I^{\mathrm{p}}$
      - Compute $\Sigma_E^{\mathrm{k}} W_{\mathrm{k}}^{\mathrm{iter}} = \Lambda_{\mathrm{k}} \Sigma_{\mathrm{I}}^{\mathrm{k}} W_{\mathrm{k}}^{\mathrm{iter}}$, obtain $W_{\mathrm{k}}^{\mathrm{iter}}$.
   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. Sort the $\acute{\mathrm{I}}_{\mathrm{N}}$ eigenvectors $W_{\mathrm{k}}^{\mathrm{iter}} \in \Re^{\mathrm{I_k} \times \acute{\mathrm{I}}_{\mathrm{k}}}$ according to Λ_{k} in decreasing order, $\mathrm{k} = 1, \cdots, \mathrm{N}$.

6.3 Proposed Tensor Kinship verification pipeline

In this section, we explain the details of employing the proposed TXQDA for kinship verification from pairs of face images. As depicted in Fig. 6.1, the block diagram of the proposed approach consists of three essential components: feature extraction, tensor subspace transformation and comparison. We focus in this work on subspace transformation and the feature extraction based multiple scales local descriptor.

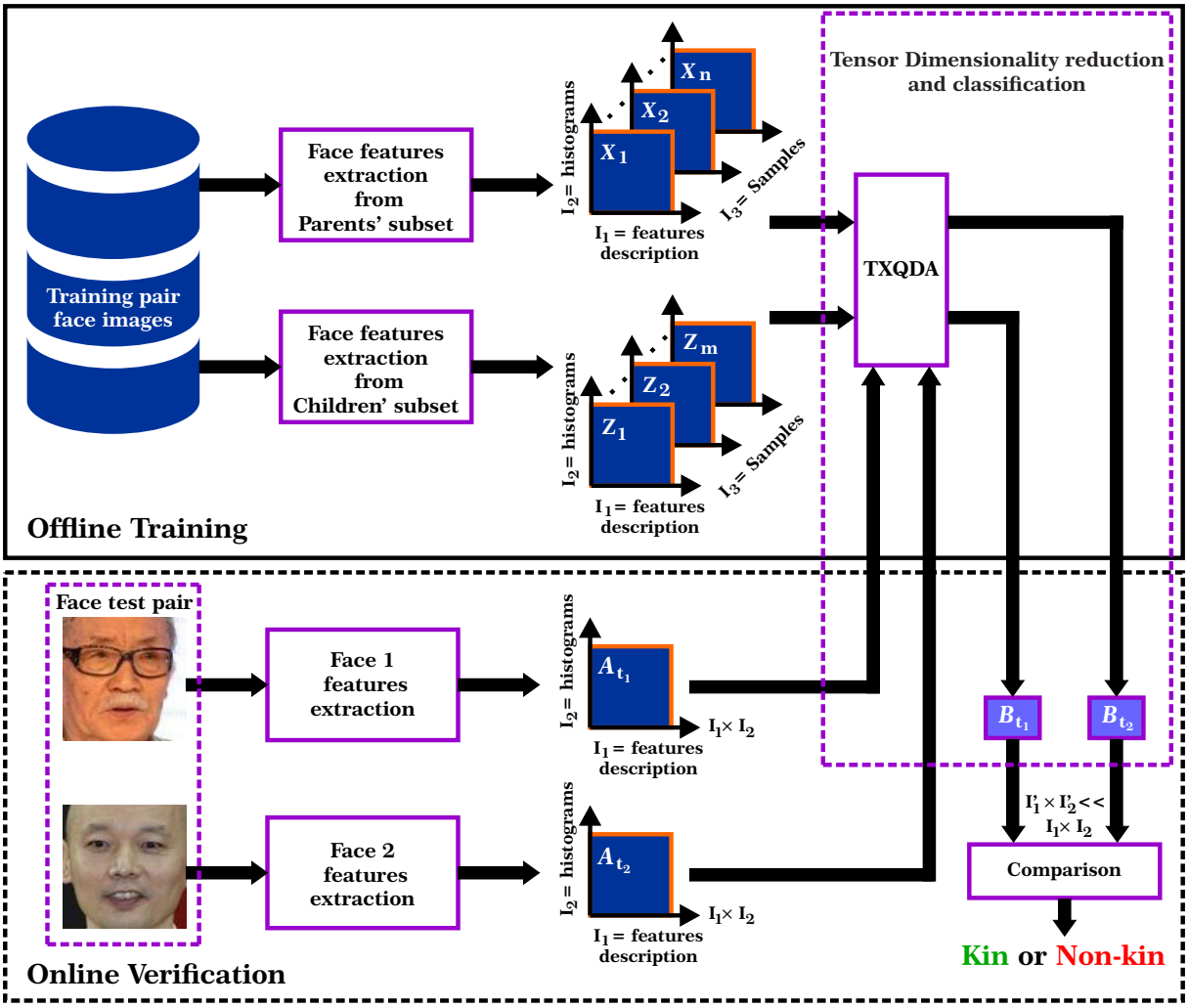


Figure 6.1: Block diagram of the proposed face pair matching system.

6.3.1 Feature extraction

To describe face images, we extract two popular local texture descriptors: the Binarized Statistical Image Feature (BSIF) [77] and the Local Phase Quantization (LPQ) [120]. To increase the verification rate, we extract the two descriptors at multiple scales by varying the values of their parameters, i.e., W the filter size of BSIF; M the window size of LPQ.

6.3.2 Tensor Design

In the offline (training) stage, the optimal multilinear projection matrices are estimated, and in online (test) stage, new samples are projected by these tensors and matched. The training 3^{rd} order tensors $\boldsymbol{X}, \boldsymbol{Z} \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ is constructed using the histograms of different local descriptors extracted from the training face images. The three modes of the tensors $\boldsymbol{X}$ and $\boldsymbol{Z}$ are defined as follows: I_1 corresponds to the local descriptors extracted at different scales, I_2 represents the histograms, and I_3 face samples in the database .

The input tensors $\boldsymbol{X}$ and $\boldsymbol{Z}$ are reduced according to I_1 and I_2 modes and projected into another subspace based on the proposed TXQDA method. Then, we obtain a reduced tensor with $\acute{\mathrm{I}_1} \times \acute{\mathrm{I}_2} \ll \mathrm{I}_1 \times \mathrm{I}_2$.

The training data of TXQDA method includes the match pairs (positive pairs) only. The two tensors ($\boldsymbol{X}$ and $\boldsymbol{Z}$) are used to compute the covariance matrix of the intrapersonal variations Σ_I and the covariance matrix of the intrapersonal variations Σ_E of the TXQDA method.

In the test phase, each face pair passes the same steps of feature extraction as in the training phase, then projected in the tensor dimensionality reduction and classification (TXQDA). Finally, the cosine similarity is used to check whether the pair of reduced features matches (belonging to the same family) or not.

6.3.3 Matching

To compare between two faces pair, we used the reduced features projected through the TXQDA space which are concatenated to form one feature vector. Then, we applied cosine similarity [116] for each pair test of the two face images, so a match score is done. After discriminant analysis methods, the use of cosine similarity distance has an advantage which comes from its connection to the Bayes decision rule [95]. Cosine similarity between two vectors $\mathbf{b}_{\mathrm{t}_1}$ and $\mathbf{b}_{\mathrm{t}_2}$ is defined as the following:

$$\cos(\mathbf{b}_{\mathrm{t}_1}, \mathbf{b}_{\mathrm{t}_2}) = \frac{\mathbf{b}_{\mathrm{t}_1}^{\mathrm{T}}.\mathbf{b}_{\mathrm{t}_2}}{\|\mathbf{b}_{\mathrm{t}_1}\|.\|\mathbf{b}_{\mathrm{t}_2}\|} \tag{6.9}$$

Where $\|.\|$ is the Euclidean norm. A high value of the produced score means a high probability that $\mathbf{b}_{\mathrm{t}_1}$ and $\mathbf{b}_{\mathrm{t}_2}$ belong to the same family.

6.4 Experiments

In this section, we perform a number of experiments to evaluate the proposed kinship verification system and investigate the strengths of the Tensor representation. Firstly, we present the benchmark databases used in our experiments. Then, we discuss the parameter settings used for kinship verification. Finally, we provide and discuss the results and compare them with those of the state of the art.

6.4.1 Parameter Settings

Our approach's performance is evaluated based on the same experimental protocol found in the literature [103, 172, 174], in which five-fold cross-validation for kin verification is performed by keeping the same number of pairs images for each fold. This protocol is considered to make sure that our results are directly comparable to the state of the art. The negative pairs for kinship are generated randomly such that each image appears only once in the test set. The number of positive pairs and negative pairs is the same in the test stage.

To mitigate the effect of face normalization and to be consistent with several previous works, including [101, 103, 105, 133, 168, 174, 175, 187, 191, 193], all feature descriptions are extracted from face images that are aligned and cropped using the position of the eyes to 64×64 pixels.

Regarding feature extraction, we use eight filters with different sizes W = {3, 5, 7, 9, 11, 13, 15, 17} in the Multi-Scale Binarized Statistical Image Features (MSBSIF). In the Multi-Scale Local Phase Quantization (MSLPQ), the window size is M = {3, 5, 7, 9, 11, 13, 15, 17} . Every face image is subdivided into 16 blocks, each of size 16×16 pixels. We use histograms of 256 bins to aggregate the local features extracted from each block.

6.4.2 Results and discussion

In this subsection, we introduce and discuss the results of the proposed approach based on third order tensor representation. Moreover, all the experiments were done for the original linear approach XQDA, which works as a baseline for evaluating the proposed TXQDA method. Furthermore, the performances of the local descriptors MSLPQ and MSBSIF are separately examined as well as their fusion. In the linear case, feature level fusion is made by concatenating vectors from different scales for each face descriptors. For the 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.

Tables 6.1, 6.2, 6.3, 6.4 and 6.5 show kinship verification accuracy from different descriptors and their fusion using the proposed TXQDA method compared with linear XQDA method on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases, respectively. We remarked that the performance is improved with variation between 5% and 9%. Moreover, the proposed TXQDA method stacked the features in the second tensor mode to provide the maximum extraction of information. Consequently, many more feature dimensions are available in TXQDA than in XQDA. Furthermore, XQDA is theoretically limited by the number of classes in the data, whereas the TXQDA is not. It is also noticeable that the best results are obtained by ($\text{MSLPQ}_{(3+5+7+9+11)}$ + $\text{MSBSIF}_{(3+5+7+9+11)}$) description.

In Table 6.3, the tri-subject kinship verification is performed by Logistic Regression

(LR) [56] score fusion method as used in [83]. The scores of Father-Son and Mother-Son are fused to generate FM-S scores for tri-subject matching. The scores of Father-Daughter and Mother-Daughter are fused to generate FM-D scores for tri-subject matching.

As shown in Table 6.2 our TXQDA method processes the age difference factor and benefits from the child-young parent set (Set 1) in which this set is gathered to lighten the age difference shown in the child-old parent set (Set 2). XQDA method does not benefit from the child-young parent set and this demonstrates that the XQDA method neglected to process the age difference factor. Moreover, the multifactor structure (kinship, gender, age, expression, illumination and pose) was analyzed and separated from different tensor dimensions of the proposed TXQDA method.

Table 6.6 shows kinship verification accuracy of the proposed TXQDA method compared with linear XQDA method and three other methods (i.e. Neighborhood Repulsed Metric Learning [103] (NRML) method, Side-Information based Linear Discriminant analysis [109] (SILD) method and Multilinear Side-Information based Discriminant Analysis [14] (MSIDA) method) using (MSLPQ+MSBSIF) features description on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases.

Moreover, the NRML and SILD methods, which are the most used methods for kinship verification, give lower performance than the XQDA and TXQDA methods in all cases. The performance becomes clear, significant and better by using the viewpoint changes methods compared with NRML and SILD methods.

Furthermore, Cornell KinFace and UB KinFace databases are gathered with mixture of four kin relations, which make hard learning of kin relations from different persons with different gender and with high significant age difference. Our proposed Tensor cross-view based method shows the superiority with a large margin in results compared with NRML and SILD methods on Cornell KinFace and UB KinFace. On Cornell KinFace database, TXQDA method performs with 9% of performance better than XQDA method, 9.42% of performance better than MSIDA, 18% of performance better than SILD and 17.5 % of performance better than NRML. On UB KinFace database, TXQDA method works with a performance of 9% better than XQDA method, a performance of 9.48% better than MSIDA, a performance of 24% better than SILD and a performance of 21 % better than NRML.

This demonstrates that the proposed Tensor XQDA works well on the difficult cases, where the viewpoint changes exist under the mixture of four kin relations (i.e. Father-Daughter and Mother-Son are a face images pairs that include two different person with different gender, and Father-Son and Mother-Daughter are a face images pairs that include two different person with same gender). Furthermore, TXQDA method works with a performance of 4.4% better than XQDA method, a performance of 4.83% better than MSIDA, a performance of 6.8% better than SILD and a performance of 7.3 % better than NRML on the TSKinFace database. Our TXQDA method works with a performance

of 5.50% better than XQDA method, a performance of 5.00% better than MSIDA, a performance of 10.65% better than SILD and a performance of 14.20% better than NRML on the KinFaceW-II database. Moreover, our TXQDA method works with a performance of 7.93% better than XQDA method, a performance of 9.98% better than MSIDA, a performance of 11.05% better than SILD and a performance of 9.61% better than NRML on the FIW database.

Besides, our results show that the multifactor structure belonging to kinship, gender, age, expression, illumination and pose is taken into consideration in our TXQDA method to a large extent than the XQDA method. It is remarkable that SILD method used the positive and negative pairs in the training stage unlike the proposed TXQDA which obtained better performances than SILD by using only the positive pairs. Our TXQDA presented the face image as a matrix, where the face feature descriptions are stacked in a tensor mode providing the maximum extraction of information. Unlike in the case of XQDA method, the feature vectors are purely concatenated neglecting the natural structure of data. As shown in our results, the tensor is an elegant way of presenting and fusing data.

Table 6.1: The mean accuracy (%) of kinship verification for TXQDA and XQDA using different MSLPQ and MSBSIF scales and their fusion on the Cornell KinFace database.

Descriptor	XQDA	TXQDA
	Mean	Mean
$MSLPQ_{(3+5+7)}$	83.93	91.71
$MSLPQ_{(5+7+9)}$	81.07	92.71
$MSLPQ_{(9+11+13)}$	77.17	90.59
$MSLPQ_{(13+15+17)}$	76.51	89.89
$MSLPQ_{(3+5+7+9)}$	83.12	92.38
$MSLPQ_{(3+5+7+9+11)}$	83.83	92.74
$MSBSIF_{(3+5+7)}$	81.61	92.66
$MSBSIF_{(5+7+9)}$	79.60	92.32
$MSBSIF_{(9+11+13)}$	77.15	92.28
$MSBSIF_{(13+15+17)}$	79.65	89.85
$MSBSIF_{(3+5+7+9)}$	80.97	92.00
$MSBSIF_{(3+5+7+9+11)}$	80.98	92.33
$MSLPQ_{(3+5+7+9+11)}$ + $MSBSIF_{(3+5+7+9+11)}$	**84.10**	**93.04**

Table 6.2: The mean accuracy (%) of kinship verification for TXQDA and XQDA using different MSLPQ and MSBSIF scales and their fusion on the UB KinFace database.

Descriptor	XQDA			TXQDA		
	Set 1	Set 2	Mean	Set 1	Set 2	Mean
$MSLPQ_{(3+5+7)}$	76.74	80.21	78.48	91.28	90.53	90.91
$MSLPQ_{(5+7+9)}$	78.50	80.96	79.73	91.03	89.27	90.15
$MSLPQ_{(9+11+13)}$	74.71	76.19	75.45	90.25	88.29	89.27
$MSLPQ_{(13+15+17)}$	75.47	75.66	75.57	87.00	88.04	87.52
$MSLPQ_{(3+5+7+9)}$	79.22	80.19	79.71	91.28	90.02	90.65
$MSLPQ_{(3+5+7+9+11)}$	78.46	80.93	79.70	91.76	90.26	91.01
$MSBSIF_{(3+5+7)}$	77.00	76.73	76.87	91.02	90.77	90.90
$MSBSIF_{(5+7+9)}$	79.69	78.90	79.30	91.26	91.28	91.27
$MSBSIF_{(9+11+13)}$	79.23	80.47	79.85	91.28	90.76	91.02
$MSBSIF_{(13+15+17)}$	77.39	73.43	75.41	89.78	89.02	89.40
$MSBSIF_{(3+5+7+9)}$	82.25	79.70	80.98	91.25	91.03	91.14
$MSBSIF_{(3+5+7+9+11)}$	79.45	80.41	79.93	91.51	90.77	91.14
$MSLPQ_{(3+5+7+9+11)}$ + $MSBSIF_{(3+5+7+9+11)}$	82.24	82.92	**82.58**	92.03	91.02	**91.53**

Table 6.3: The mean accuracy (%) of kinship verification for TXQDA and XQDA using different MSLPQ and MSBSIF scales and their fusion on the TSKinFace database.

Descriptor	XQDA							TXQDA						
	F-S	F-D	M-S	M-D	Mean	FM-S	FM-D	F-S	F-D	M-S	M-D	Mean	FM-S	FM-D
$MSLPQ_{(3+5+7)}$	86.84	83.47	83.33	85.06	84.68	87.41	85.46	87.18	88.42	88.54	88.69	88.21	92.82	94.45
$MSLPQ_{(5+7+9)}$	83.63	81.78	83.33	84.66	83.35	86.64	85.36	84.27	85.05	85.73	86.11	85.29	91.26	91.77
$MSLPQ_{(9+11+13)}$	81.48	79.18	79.53	81.37	80.39	85.20	84.86	82.23	81.78	84.08	83.13	82.81	90.58	90.18
$MSLPQ_{(13+15+17)}$	80.90	78.69	78.95	78.28	79.21	85.38	83.67	81.55	80.89	83.69	82.63	82.19	90.19	87.89
$MSLPQ_{(3+5+7+9)}$	85.77	84.66	84.99	86.35	85.44	87.31	86.15	87.09	87.72	88.16	89.09	88.02	92.91	94.45
$MSLPQ_{(3+5+7+9+11)}$	86.84	84.06	83.72	84.76	84.85	87.22	85.95	88.35	88.61	88.25	89.28	88.62	93.01	94.44
$MSBSIF_{(3+5+7)}$	83.14	82.76	81.77	85.15	83.21	85.78	85.25	88.06	88.32	88.54	89.68	88.65	93.98	94.34
$MSBSIF_{(5+7+9)}$	84.32	82.57	81.97	84.56	83.36	86.74	86.85	86.41	88.12	88.54	88.10	87.79	93.69	94.74
$MSBSIF_{(9+11+13)}$	82.95	79.99	80.51	82.67	81.53	84.79	84.36	85.15	85.45	86.80	86.01	85.85	92.52	91.96
$MSBSIF_{(13+15+17)}$	82.27	75.90	79.14	79.29	79.15	84.99	85.45	84.17	83.37	86.70	85.11	84.84	91.75	91.46
$MSBSIF_{(3+5+7+9)}$	84.90	84.26	82.26	86.15	84.39	87.52	87.35	89.03	88.32	89.13	89.48	88.99	94.08	94.44
$MSBSIF_{(3+5+7+9+11)}$	85.29	83.06	84.02	86.35	84.68	87.32	87.35	89.03	88.12	89.13	89.38	88.92	94.08	94.34
$MSLPQ_{(3+5+7+9+11)}$ + $MSBSIF_{(3+5+7+9+11)}$	87.04	85.26	84.40	86.85	**85.89**	**88.02**	**88.35**	89.32	90.69	90.29	90.97	**90.32**	**94.85**	**95.63**

Table 6.4: The mean accuracy (%) of kinship verification for TXQDA and XQDA using different MSLPQ and MSBSIF scales and their fusion on the KinFaceW-II database.

Descriptor	XQDA					TXQDA				
	F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean
$MSLPQ_{(3+5+7)}$	83.80	77.40	79.00	80.20	80.10	88.00	83.20	83.60	83.60	84.60
$MSLPQ_{(5+7+9)}$	83.80	78.20	79.60	77.80	79.85	89.20	83.40	84.00	84.20	85.20
$MSLPQ_{(9+11+13)}$	81.80	79.00	77.80	78.40	79.25	87.00	81.40	83.80	85.20	84.35
$MSLPQ_{(13+15+17)}$	80.80	77.60	75.40	78.20	78.00	89.00	80.80	81.80	83.60	83.80
$MSLPQ_{(3+5+7+9)}$	84.20	77.60	80.20	80.00	80.50	89.20	83.80	83.00	84.80	85.20
$MSLPQ_{(3+5+7+9+11)}$	84.00	78.80	80.40	80.00	80.80	89.40	83.40	83.60	85.00	85.35
$MSBSIF_{(3+5+7)}$	83.60	78.60	79.20	79.60	80.25	86.60	85.20	82.20	83.20	84.30
$MSBSIF_{(5+7+9)}$	84.40	79.40	78.80	77.80	80.10	88.00	84.20	83.40	82.00	84.10
$MSBSIF_{(9+11+13)}$	83.40	78.60	77.40	77.20	79.15	86.60	82.20	81.20	82.00	83.00
$MSBSIF_{(13+15+17)}$	82.00	76.00	77.20	76.60	77.95	86.40	81.20	82.80	81.60	83.00
$MSBSIF_{(3+5+7+9)}$	84.40	80.20	79.20	79.60	80.85	87.60	85.20	81.60	83.80	84.55
$MSBSIF_{(3+5+7+9+11)}$	84.60	79.20	78.80	79.60	80.55	88.00	85.80	82.20	85.00	85.25
$MSLPQ_{(3+5+7+9+11)}$ + $MSBSIF_{(3+5+7+9+11)}$	85.00	80.60	80.60	80.40	**81.65**	90.20	86.40	85.60	86.40	**87.15**

Table 6.5: The mean accuracy (%) of kinship verification for TXQDA and XQDA using different MSLPQ and MSBSIF scales and their fusion on the four grandparent-grandchild subsets of FIW database.

Descriptor	XQDA					TXQDA				
	GF-GD	GF-GS	GM-GD	GM-GS	Mean	GF-GD	GF-GS	GM-GD	GM-GS	Mean
$MSLPQ_{(3+5+7)}$	55.91	57.92	58.08	57.59	57.38	65.57	63.92	63.85	65.08	64.61
$MSLPQ_{(5+7+9)}$	55.73	58.07	57.78	57.87	57.36	65.54	64.81	63.96	64.62	64.73
$MSLPQ_{(9+11+13)}$	55.98	58.30	57.85	57.16	57.32	66.11	64.23	63.00	64.54	64.47
$MSLPQ_{(13+15+17)}$	56.07	58.06	57.22	56.70	57.01	65.49	65.18	62.65	64.14	64.37
$MSLPQ_{(3+5+7+9)}$	56.15	58.29	57.98	58.54	57.74	65.89	64.43	63.82	64.71	64.71
$MSLPQ_{(3+5+7+9+11)}$	56.36	58.90	58.27	58.15	57.92	65.32	63.79	63.49	64.15	64.19
$MSBSIF_{(3+5+7)}$	55.85	59.06	58.46	57.26	57.66	65.35	64.50	64.32	64.61	64.70
$MSBSIF_{(5+7+9)}$	55.58	58.03	59.27	57.06	57.49	65.25	64.21	63.52	64.45	64.36
$MSBSIF_{(9+11+13)}$	56.30	56.80	58.44	56.73	57.07	66.00	65.04	64.74	64.95	65.18
$MSBSIF_{(13+15+17)}$	56.32	58.01	57.81	56.53	57.17	65.46	64.18	63.24	64.75	64.41
$MSBSIF_{(3+5+7+9)}$	56.04	58.61	59.00	57.60	57.81	65.20	64.59	63.15	64.43	64.34
$MSBSIF_{(3+5+7+9+11)}$	55.93	58.43	59.35	57.70	57.85	65.26	64.22	63.78	65.33	64.65
$MSLPQ_{(3+5+7+9+11)}$ + $MSBSIF_{(3+5+7+9+11)}$	56.04	59.23	59.00	58.12	**58.10**	66.43	66.79	65.24	65.67	**66.03**

Table 6.6: Comparison verification accuracy (%) of the proposed TXQDA with XQDA, NRML, SILD and MSIDA methods using MSLPQ+MSBSIF features description on the Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases.

Method	Cornell KinFace	UB KinFace	TSKinFace			KinFaceW-II	FIW
	Mean	Mean	Mean	FM-S	FM-D	Mean	Mean
NRML	75.52	70.55	80.83	84.26	85.53	72.95	56.42
SILD	71.38	67.36	83.46	86.44	87.82	76.50	54.98
XQDA	84.10	82.58	85.89	88.02	88.35	81.65	58.10
MSIDA	83.62	82.05	85.49	91.26	91.27	82.15	56.05
TXQDA	**93.04**	**91.53**	**90.32**	**94.85**	**95.63**	**87.15**	**66.03**

To analyze the performance of different kinship relations, we plot in Figures 6.2, 6.3, 6.4 and 6.5 the ROC curves of different methods (NRML, SILD, XQDA, MSIDA and TXQDA) using the best performing features (MSLPQ+MSBSIF) on UB KinFace set 1, UB KinFace set 2, Cornell KinFace, TSKinFace, KinFaceW-II and FIW databases, respectively.

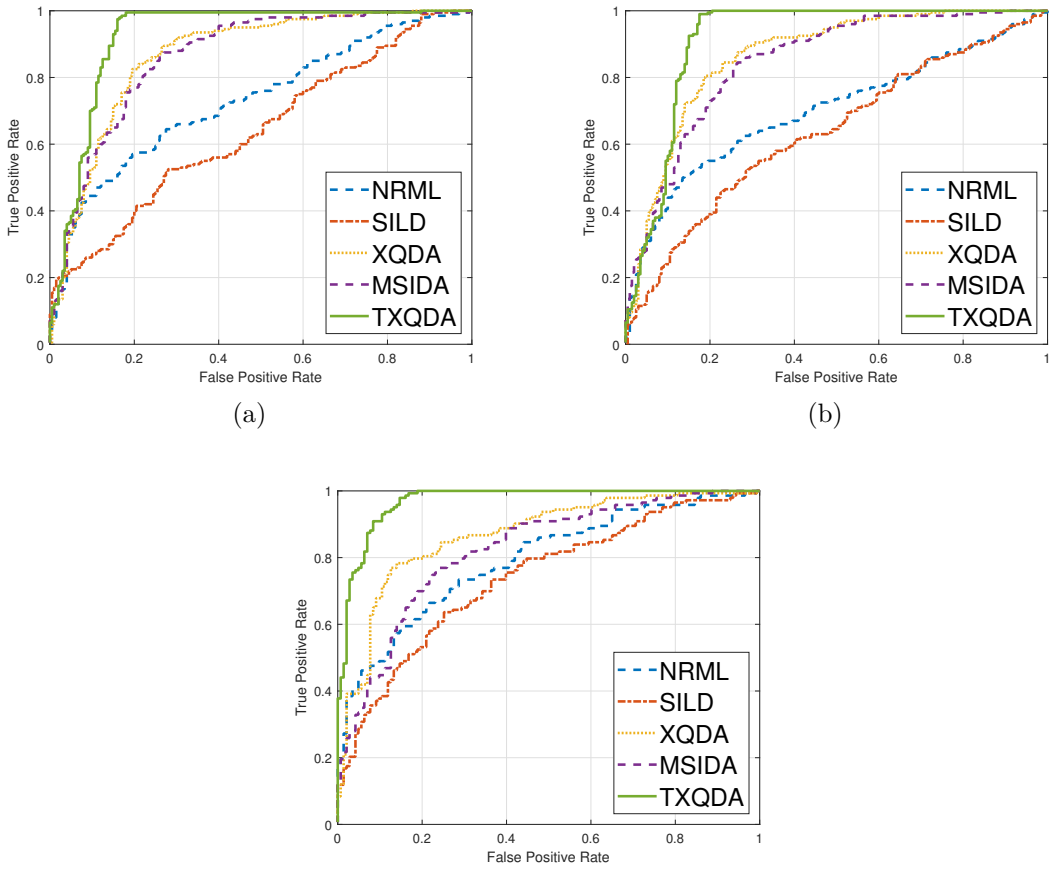


Figure 6.2: ROC curves of different methods (NRML, SILD, XQDA, MSIDA and TXQDA) using the best performing features (MSLPQ+MSBSIF) obtained on (a) UB KinFace set 1, (b) UB KinFace set 2, (c) Cornell KinFace databases, respectively.

6.4.3 The robustness's evaluation of the proposed TXQDA method

In this subsection, we tested the robustness of the proposed tensor method on TSKinFace database through additive noise and degradation of test set. For the clarification, we express the interference of face recognition $\eta = \eta_f + \eta_q$ [48], where η_f indicates facial variations such as kinship, expression, illumination, misalignment and age, and η_q indicates the image variation due to sensor or coding-related issues, such as Gaussian noise, blur, compression, and low resolution. Most of the studies on the TSKinFace database focused only on the effect of η_f, whereas our extended experiments study both the pure effect of η_q and the superposed interference of $\eta_f + \eta_q$. For an inclusive study, we generate four types of noise or degradations that are most common in real-world systems but that have not appeared in the standard database. Specifically, we generate the following versions of test sets: 1) three levels of Gaussian noise. The images are normalized in the range of (0; 1), and then we apply additive Gaussian noise with zero mean and standard derivations of σ = 0.01; 0.02; 0.03; 2) three different Gaussian blur test sets using a Gaussian kernel of size 10×10 with $\sigma = \{1; 3/2; 2\}$; 3) three different compressed images using MATLAB's JPEG

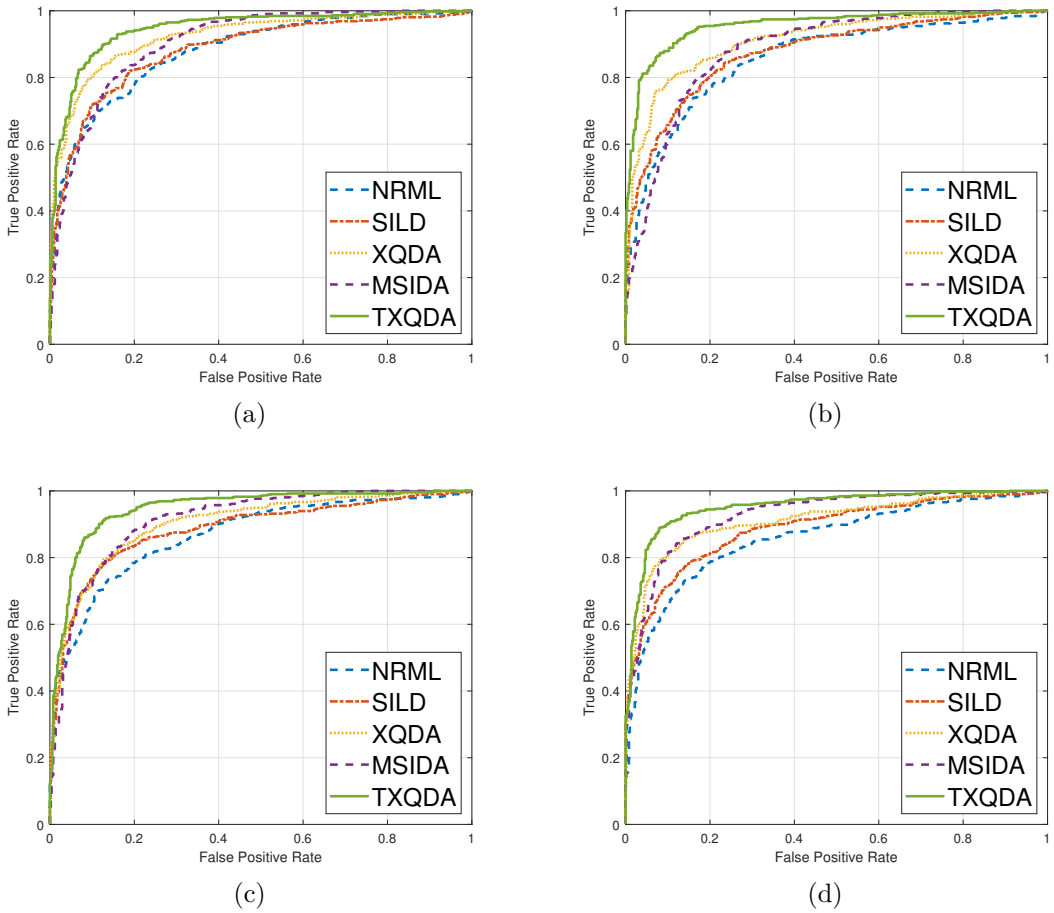


Figure 6.3: ROC curves of different methods (NRML, SILD, XQDA, MSIDA and TXQDA) using the best performing features (MSLPQ+MSBSIF) on TSKinFace database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

codec of quality 60, 45 and 30; and 4) three different low-resolution sets of test images by first downsampling the images by ratios of 2, 3, and 4 and then interpolating them to the original resolution by the "nearest" method in MATLAB. Example test images are shown in Fig. 6.6, and as shown in this figure, these degraded faces are recognizable by humans and are very common in real-world surveillance scenarios. Therefore, it is important to study how the accuracy of the metric learning methods change under these degradations. Table 6.7 shows that TXQDA and XQDA show much better robustness than the other methods under image blur, noise, compression, and reduced resolution. However, our Tensor XQDA preserves the data structure and extracts more discriminative information from degraded face images compared with XQDA method.

6.4.4 Computational cost

We computed the computational time needed for the kinship verification of one pair face samples (Parent-Child) using different methods. The experiments were implemented using MATLAB 2014a on a PC with an Intel Core i7 2.00 GHz CPU and 8 GB of RAM. The

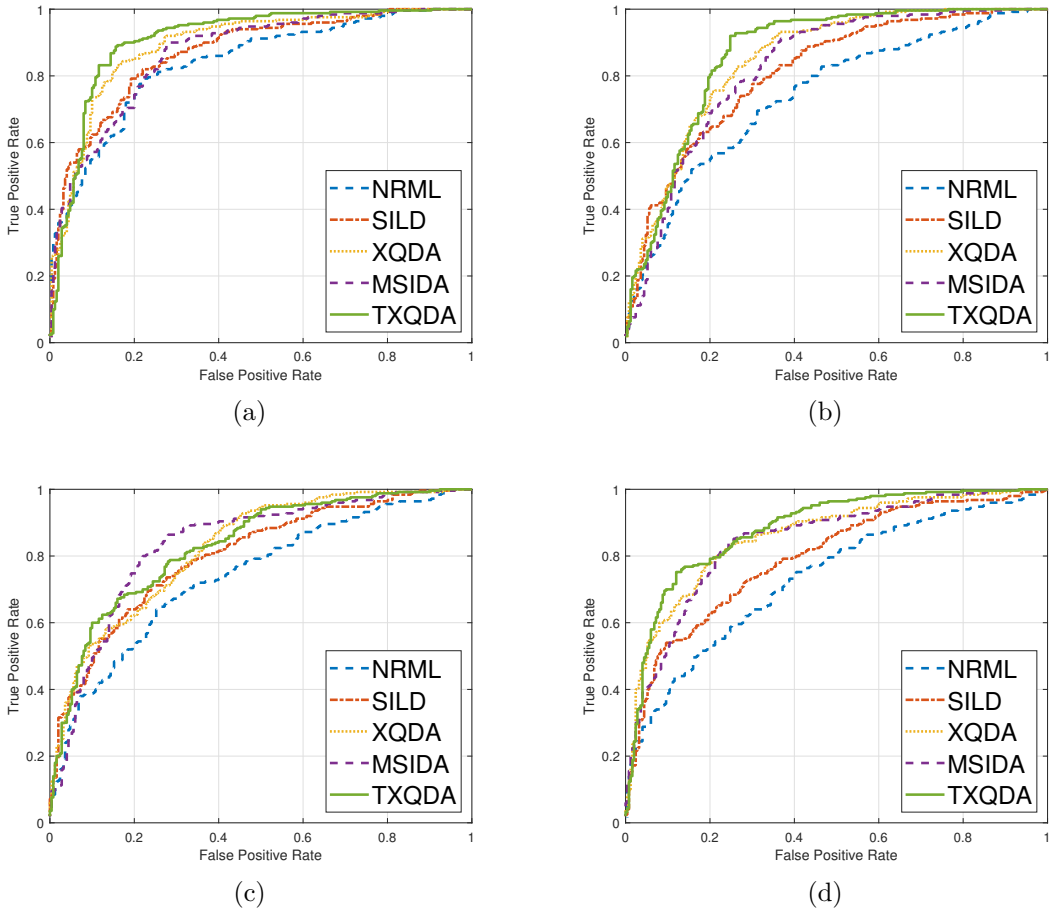


Figure 6.4: ROC curves of different methods (NRML, SILD, XQDA, MSIDA and TXQDA) using the best performing features (MSLPQ+MSBSIF) on KinFaceW-II database obtained on (a) F-S set, (b) F-D set, (c) M-S set and (d) M-D set, respectively.

Table 6.7: Comparative verification rates (%) of extended TSKinFace evaluation on the robustness to the four types of common degradations. Accuracy loss of each degradation degree on each test set is reported in detail.

Method	Relation Type	Basic Accuracy [1]	Gaussian Blur			Gaussian Noise			JPEG Compression			Reduced Resolution			Summarized Accuracy [2]
			1	3/2	2	0.01	0.02	0.03	60	45	30	1/2	1/3	1/4	
NRML	FS	81.00	-7.02	-7.81	-12.68	-6.93	-8.49	-14.33	-5.95	-6.63	-7.22	-6.73	-7.12	-12.97	72.34 (-8.66)
	FD	79.88	-6.27	-8.47	-13.65	-7.67	-9.66	-13.45	-7.07	-9.16	-12.85	-6.57	-8.67	-13.94	70.09 (-9.79)
	MS	81.48	-8.67	-10.92	-14.13	-9.25	-11.11	-13.15	-9.06	-10.62	-13.64	-8.48	-10.43	-13.94	70.36 (-11.12)
	MD	80.97	-5.38	-9.37	-13.85	-5.38	-8.47	-13.35	-5.38	-8.67	-13.15	-5.38	-8.67	-13.95	71.72 (-9.25)
	Mean	80.83	-6.83	-9.14	-13.57	-7.3	-9.43	-13.57	-6.86	-8.77	-11.71	-6.79	-8.72	-13.7	71.13 (-9.7)
SILD	FS	84.60	-9.25	-10.81	-12.76	-9.35	-10.23	-13.35	-6.43	-8.57	-9.06	-9.06	-9.45	-11.20	74.64 (-9.96)
	FD	81.76	-9.05	-11.64	-12.94	-11.64	-11.84	-12.44	-8.65	-8.85	-9.25	-8.65	-9.35	-10.94	71.32 (-10.44)
	MS	83.72	-10.52	-12.86	-16.46	-10.71	-12.57	-15.68	-10.32	-11.01	-11.20	-11.30	-11.98	-13.44	71.38 (-12.34)
	MD	83.76	-9.17	-11.16	-15.04	-8.88	-10.66	-14.55	-9.97	-11.06	-14.15	-10.67	-10.87	-11.56	72.28 (-11.48)
	Mean	83.46	-9.50	-11.62	-14.30	-10.15	-11.32	-14.01	-8.84	-9.87	-10.91	-9.92	-10.41	-11.79	72.41 (-11.05)
XQDA	FS	87.04	-6.73	-8.87	-10.72	-5.85	-7.12	-8.87	-3.90	-7.12	-7.99	-6.14	-8.87	-10.73	79.30 (-7.74)
	FD	85.26	-5.08	-7.77	-10.25	-4.08	-4.98	-7.07	-3.59	-4.78	-4.98	-3.98	-6.68	-10.16	79.14 (-6.12)
	MS	84.40	-4.09	-6.05	-8.68	-3.61	-4.48	-5.75	-3.12	-4.87	-5.27	-3.61	-5.65	-8.87	79.06 (-5.34)
	MD	86.85	-5.88	-7.27	-8.87	-5.68	-7.87	-7.97	-5.68	-7.27	-7.27	-6.28	-6.87	-9.16	79.68 (-7.17)
	Mean	85.89	-5.45	-7.49	-9.63	-4.81	-6.11	-7.42	-4.08	-6.01	-6.38	-5.00	-7.02	-9.73	79.30 (-6.59)
TXQDA	FS	89.32	-1.46	-4.56	-7.18	-0.87	-4.37	-6.80	-2.14	-3.01	-3.59	-2.62	-4.37	-7.09	85.32 (-4.00)
	FD	90.69	-2.27	-5.15	-9.70	-1.38	-4.95	-8.41	-3.07	-4.16	-4.55	-3.26	-5.74	-10.49	85.43 (-5.26)
	MS	90.29	-1.94	-5.63	-8.64	-0.78	-4.46	-7.57	-1.26	-3.69	-4.17	-2.62	-4.66	-7.86	85.85 (-4.44)
	MD	90.97	-1.68	-5.46	-8.53	-1.29	-5.65	-6.95	-1.59	-4.17	-5.46	-2.48	-4.96	-8.13	86.27 (-4.70)
	Mean	90.32	-1.84	-5.20	-8.51	-1.08	-4.86	-7.43	-2.02	-3.76	-4.62	-2.75	-4.93	-8.39	**85.72 (-4.60)**

To provide a comprehensive result, the verification accuracy across the three types of probe sets is reported.

[1] The verification accuracy on the original TSKinFace database.

[2] The verification accuracy across all types and all degrees of the tested noise and degradations.

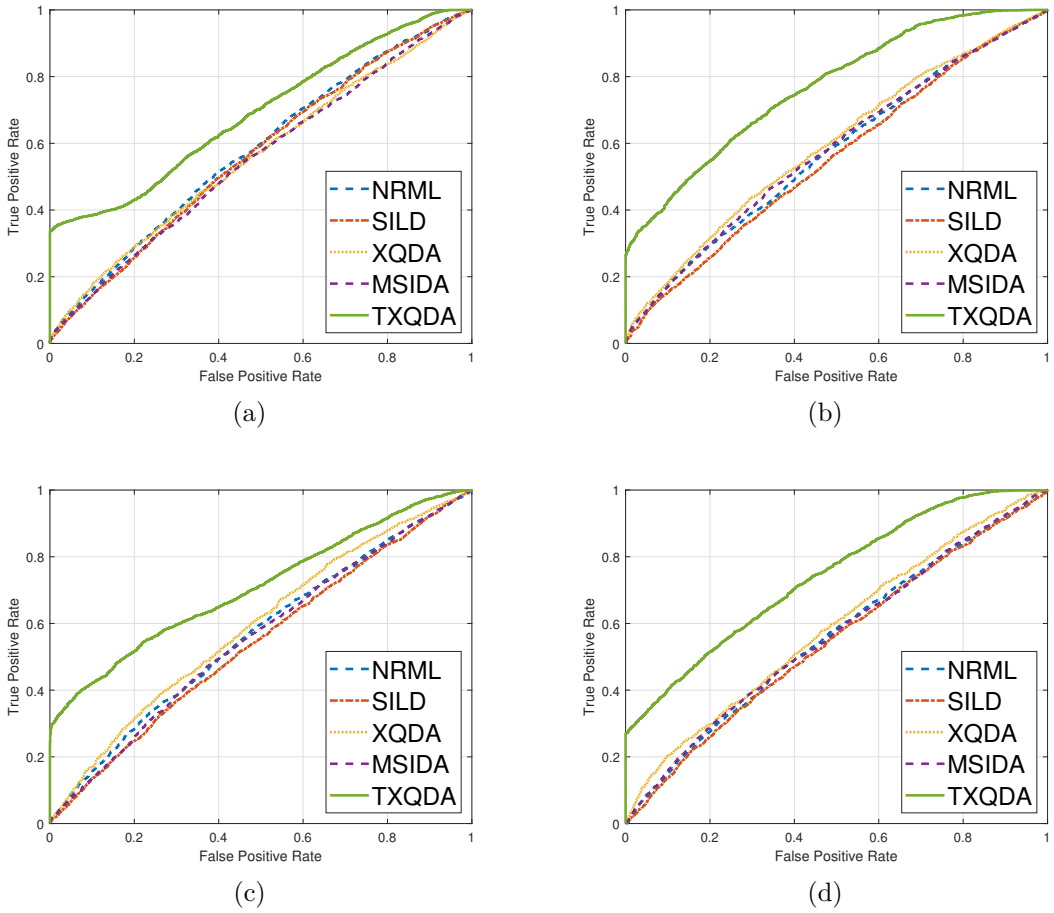


Figure 6.5: ROC curves of different methods (NRML, SILD, XQDA, MSIDA and TXQDA) using the best performing features (MSLPQ+MSBSIF) on FIW database obtained on (a) GF-GD set, (b) GF-GS set, (c) GM-GD set and (d) GM-GS set, respectively.

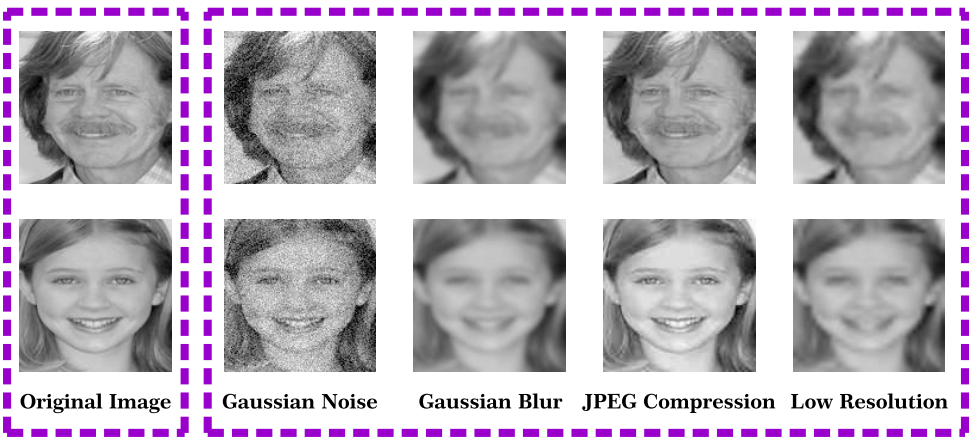


Figure 6.6: Examples of original and degraded images used in our extended TSKinFace evaluation. The last four columns correspond to the most severe degrees of Gaussian noise, Gaussian blur, JPEG compression, and reduced resolution applied on the test images.

feature extraction for the case of MSLPQ+MSBSIF description takes 0.104 s. In training stage (offline), the estimation of the projection matrices is performed only once. In the online phase, we evaluate the time cost needed by each method to project and match the test pair which is provided in Table 6.8 in ms. This table shows that the best performing variant, TXQDA, runs faster compared with all the other methods. Furthermore, we see that the time cost of projection and matching is negligible compared to the feature extraction time. The total time cost for our framework using MSLPQ+MSBSIF features and TXQDA method is about 0.109 s for both Cornell KinFace and UB KinFace databases, and 0.110 s from TSKinFace and KinFaceW-II databases, and 0.115 s for FIW database.

Table 6.8: Time Cost (TC), in ms, taken by different methods for the projection of one pair of face images.

Database	Feature extraction	Projection and matching					All steps				
		NRML	SILD	XQDA	MSIDA	TXQDA	NRML	SILD	XQDA	MSIDA	TXQDA
Cornell KinFace	104.92	14.22	22.93	10.61	5.22	**4.54**	119.14	127.85	115.53	110.14	**109.46**
UB KinFace		15.17	11.91	7.06	5.73	**4.94**	120.09	116.83	111.98	110.65	**109.86**
TSKinFace		43.04	12.15	10.92	5.90	**5.88**	147.96	117.07	115.84	110.82	**110.80**
KinFaceW-II		21.17	11.99	11.03	6.32	**6.05**	126.09	116.91	115.95	111.24	**110.97**
FIW		97.27	90.11	78.96	21.16	**10.41**	102.19	195.03	183.88	126.08	**115.33**

6.4.5 Comparison with the results of the state of the art

The best kinship verification performances of our approach are achieved using two descriptors ($\text{MSLPQ}_{(3+5+7+9+11)}$ + $\text{MSBSIF}_{(3+5+7+9+11)}$) on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases. For the XQDA (linear), verification rates of 84.10%, 82.58%, 85.89%, 81.65% and 58.10% are reported on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases, respectively. For TXQDA (multilinear), verification rates of 93.04%, 91.53%, 90.32%, 87.15% and 66.03% are reported on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases, respectively. These results are compared with the state of the art in Tables 6.9, 6.10, 6.11, 6.12 and 6.13. The comparison reveals that the proposed Tensor cross-view analysis based method outperforms the recent state of the art on the five databases, Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW. This demonstrates the effectiveness of using the cross-view methods in kinship verification topic.

TXQDA vs. KVRL-fcDBN [79]: In this comparison, we focus on two databases Cornell KinFace and UB KinFace. The work of Kohli *et al.* [79] was based on a deep learning approach (KVRL-fcDBN), where fcDBN algorithm was used to learn more than 600,000 outside face images and obtained 89.50 % on Cornell KinFace and 91.80 % on UB KinFace databases. However, in our work, we used Tensor cross-view metric learning method (i.e. TXQDA) to learn the provided data only (i.e. no outside data was used) and we achieved good performances, 93.04 % and 91.53 % on Cornell KinFace and UB KinFace databases, respectively. Our TXQDA outperforms the KVRL-fcDBN approach on Cornell KinFace database and got a competitive performance on UB KinFace database.

TXQDA vs. DDML [101], DDMML [101] and MvDML [60]: In this comparison, we focus on TSKinFace and KinFaceW-II databases for monocular and multiple feature description. From mono-view feature description, the Discriminative Deep Metric Learning (DDML) [101] method used LPQ descriptor with face images with input size of 64×64. First they divide each face image into 4×4 non-overlapping blocks, where the size of each block is 16×16. Then, they extract a 256-bin LPQ histogram with window size of 3, 5 and 7 for each block respectively, and finally concatenate them to form a 12,288-D feature vector, which is the same used face features description in our work (i.e. $\text{MSLPQ}_{(3+5+7)}$). For bi-subject kinship verification, from the $\text{MSLPQ}_{(3+5+7)}$ features description, our TXQDA method outperforms the DDML method with about 8.28 % and 2.40 % on TSKinFace and KinFaceW-II databases, respectively. For tri-subject kinship verification, from the $\text{MSLPQ}_{(3+5+7)}$ features description, our TXQDA method outperforms the DDML method with about 9.12 % and 11.25 % on TSKinFace database for FM-S and FM-D relations, respectively. From multi-view feature description, we can see that our approach which using the MSLPQ+MSBSIF feature description, improves the performance with about 2.85 % and 6.95 % compared with DDMML and MvDML on KinFaceW-II database, respectively. Furthermore, our TXQDA method improves the performances with about 6 % compared with DDMML for bi-subject kinship verification on TSKinFace database and improves the performances with about 6.35 % and 8.53 % compared with DDMML for FM-S and FM-D tri-subject relations on TSKinFace database.

TXQDA vs. MSIDA [14]: From the five kinship databases, our TXQDA outperforms the MSIDA [14] method with about 9.42%, 9.48%, 4.83%, 5.00% and 9.98% on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases, respectively. Furthermore, our TXQDA used only the positive pairs in the training step, unlike MSIDA do. The framework of Bessaoudi et al. [14] needs the MPCA method step for the features dimension reduction before using the MSIDA method. Our TXQDA method deals with the face images directly, without the need of using the features dimension reduction step. Furthermore, our TXQDA method work well compared with MSIDA method when the data classes contain face images from different persons with very large age difference (i.e. the four grandparent-grandchild relations of FIW database).

Table 6.9: Performance comparisons (%) with state-of-the-art methods on Cornell KinFace database.

Method	Mean Accuracy (%)
Pictorial structure model [44]	70.67
Neighborhood repulsed metric learning [103]	69.50
Multiview neighborhood repulsed metric learning [103]	71.60
Discriminative multimetric learning [174]	73.50
Prototype discriminative feature learning [175]	71.90
MHDL3 - {HOG + Color + LPQ} [105]	76.60
Multiple kernel similarity metric [193]	81.70
Heterogeneous similarity learning [131]	68.40
Kinship metric learning [195]	81.40
Multilinear side-information based discriminant analysis [14]	86.59
Neighborhood repulsed metric learning [103] (Our)	75.52
Side-information based linear discriminant analysis [109] (Our)	71.38
Cross-view quadratic discriminant analysis [93] (Our)	84.10
Multilinear side-information based discriminant analysis [14] (Our)	83.62
TXQDA (Our)	**93.04**

Table 6.10: Table 8: Performance comparisons (%) with state-of-the-art methods on UB KinFace database.

Method	Mean Accuracy (%)
Transfer subspace learning [170]	68.50
Neighborhood repulsed metric learning [103]	65.60
Multiview neighborhood repulsed metric learning [103]	67.05
Discriminative multimetric learning [174]	72.25
Prototype discriminative feature learning [175]	67.30
Heterogeneous similarity learning [131]	56.20
PML-COV-S [114]	84.50
Kinship metric learning [195]	75.50
Multilinear side-information based discriminant analysis [14]	83.34
Neighborhood repulsed metric learning [103] (Our)	70.55
Side-information based linear discriminant analysis [109] (Our)	67.36
Cross-view quadratic discriminant analysis [93] (Our)	82.58
Multilinear side-information based discriminant analysis [14] (Our)	82.05
TXQDA (Our)	**91.53**

Table 6.11: Performance comparisons (%) with state-of-the-art methods on TSKinFace database.

Method	Mean Accuracy (%)	FM-S (%)	FM-D (%)
Relative symmetric bilinear model [133]	81.85	86.40	84.40
BSIF-HSV [168]	81.19	/	/
Discriminative deep multi-metric learning [101]	84.15	88.50	87.10
Multiple kernel similarity metric [193]	84.52	/	/
Multilinear side-information based discriminant analysis [14]	85.18	/	/
SILD+WCCN/LR [83]	88.59	90.94	91.23
Neighborhood repulsed metric learning [103] (Our)	80.83	84.26	85.53
Side-information based linear discriminant analysis [109] (Our)	83.46	86.44	87.82
Cross-view quadratic discriminant analysis [93] (Our)	85.89	88.02	88.35
Multilinear side-information based discriminant analysis [14] (Our)	85.49	91.26	91.27
TXQDA (Our)	**90.32**	**94.85**	**95.63**

Table 6.12: Performance comparisons (%) with state-of-the-art methods on KinFaceW-II database.

Method	Mean Accuracy (%)
Multi-view multi-task learning [134]	77.20
Discriminative deep multi-metric learning [101]	84.30
Multi-view deep metric learning [60]	80.20
Heterogeneous similarity learning [131]	70.40
Multiple kernel similarity metric [193]	84.30
Large-margin multi-metric learning [61]	80.00
Kinship metric learning [195]	85.70
SILD+WCCN/LR [83]	86.20
Neighborhood repulsed metric learning [103] (Our)	72.95
Side-information based linear discriminant analysis [109] (Our)	76.50
Cross-view quadratic discriminant analysis [93] (Our)	81.65
Multilinear side-information based discriminant analysis [14] (Our)	82.15
TXQDA (Our)	**87.15**

Table 6.13: Performance comparisons (%) with state-of-the-art methods on the four grandparent-grandchild relations from FIW database.

Method	Mean Accuracy (%)
ResNet+CF [138]	65.51
SphereFace [138]	65.60
ResNet+SDMLoss [155]	65.58
Neighborhood repulsed metric learning [103] (Our)	56.42
Side-information based linear discriminant analysis [109] (Our)	54.98
Cross-view quadratic discriminant analysis [93] (Our)	58.10
Multilinear side-information based discriminant analysis [14] (Our)	56.05
TXQDA (Our)	**66.03**

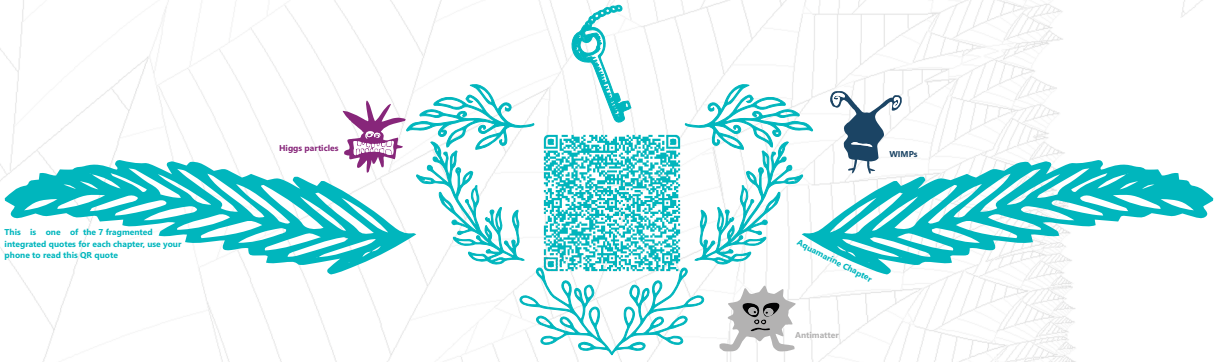
6.5 Conclusion

In this chapter, we presented an effective approach based on tensor cross-view method to the problem of kinship verification. To achieve a low dimensional and discriminative tensor subspace, we extended XQDA to TXQDA, which operate on multilinear data. TXQDA finds multilinear projections of the tensor, where the separation between data classes is enhanced. Furthermore, TXQDA was proposed to separate the multifactor structure of face images related to kinship, age, gender, expression, illumination and pose from different dimensions of the tensor. Therefore, TXQDA has many advantages as it, i) preserves data structure, ii) enlarges the margin between samples, iii) helps lightening the small sample size problem, and iv) reduces the computational cost. The experimental evaluation showed the superiority of our method. The best results of our approach are obtained by fusing histograms of two multiple scale local texture descriptors (MSLPQ+MSBSIF) projected with the proposed TXQDA method. These results outperform the state of the art on Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW databases. Furthermore, these results point out to the need of using cross-view methods for kinship

verification. As future work, we plan to investigate higher tensor orders (> 3) for face representation with the proposed multilinear dimensionality reduction method.

7 General conclusion and perspectives

Contents



7.1 General conclusion

Kinship verification from facial images, one of the new topics in computer vision that has been studied and used for several years. Moreover, facial kinship verification can be applied to potential applications, such as the creation of family trees, family album organization, image annotation, finding missing children and forensics. Checking if two persons are from the same family or not can be automatically verified through facial images. Furthermore, kinship verification shows more and more development. In addition to the bi-subject matching challenges, verifying kinship encountering the more challenging tri-subject matching problem. Learn and extract the face similarities between family members is challenging. Many encouraging results have been shown over the past a few years, kinship verification from facial images still remains open.

Over the recent years, many methods have been proposed to improve the performance of the kinship verification frameworks from facial images. Metric learning methods which are considered as one of the greatest applied dimensionality reduction methods, are widely used for face recognition. The aim of these methods is to learn and reduce the subspace of the large features in a low and distinct subspace, leading to a better separation between the data classes.

One of the most important methods, are the linear dimensionality reduction methods that include the PCA and LDA methods. The goal of PCA is to increase the variance of transformed features, extracted in the projected subspace. However, LDA maximizes the covariance of the inter-class while minimizes the covariance of the intra-class in the projected subspace. For vector-based features representation, the key shortcoming is that this representation forfeits part of the natural spatial structure of the facial images, which impedes the subsequent algorithm to erect the optimal face model and thus a weak classification model. To solve the aforementioned difficulty, many researchers proposed to use tensor representation instead of vector-based representation as the input data. Many researches proved that high order data representation based on tensor analysis, lead to increase the performances in the face recognition. The linear dimensionality reduction methods have been extended into multilinear subspace methods which are based on the high order tensor representation replaced by their vectorized forms. Furthermore, the multilinear transformations methods have the number of dimensions (N) that defines its order, analyze the multifactor structure of the face images (i.e. expression, illumination and pose) from different dimensions of the tensor. Therefore, this thesis was devoted to the problem of automatic kinship verification from faces (bi-subject and tri-subject kinship verification) by vector-based and tensor-based subspaces analysis. Our work has contributed to the development of linear and multilinear metric learning subspaces algorithms using multiple features (i.e. different scales of texture descriptors, multiple deep and shallow features and multiple shallow features). Furthermore, the combination

of multiple features was performed by two criterions as it: i) over a tensor design (early tensorial features fusion); and ii) over a score fusion (late features fusion).

In chapter 1, we gave a general introduction of the contexts, motivations, objectives and contributions of this thesis.

In chapter 2, we mentioned a general overview state of the art methods of the kinship verification as well as their different types: features learning-based kinship verification, metric learning-based kinship verification and convolutional deep learning-based kinship verification. On the other hand, we presented the kinship problem from facial images and its measuring characteristics as well as the general kinship verification system.

In chapter 3, we presented a new efficient method for facial kinship verification based on multiple scales texture feature projected through Side Information Exponential Discriminant Analysis (SIEDA) nonlinear subspace and fused several features using Logistic Regression (LR) scores fusion. Furthermore, we investigate the effectiveness of the color-texture information over discriminative subspace utilizing the two-step learning strategy, SIEDA and Logistic Regression methods, for automatic facial verification of kinship. Moreover, we tested various color spaces and descriptors on four benchmark kinship databases. Especially, each color channel of facial image from each unique color space is projected over the same implicit learned color channel subspace, and then all the channels information are combined to achieve a better discrimination.

In chapter 4, we proposed a novel discriminative subspace by extending Side-Information based Linear Discriminant analysis (SILD) to SILD integrating Within Class Covariance Normalization (SILD+WCCN) subspace transformation approach for facial kinship verification. Therfore, WCCN minimizes the class intra-variability influence by decreasing the expected classification error on the training step. Moreover, we propose two effective automated facial kinship verification frameworks appropriate for bi-subject and tri-subject matching kinship verification, from facial images captured in unconstrained environments. The facial data is appeared as a multiple feature based on the integration of various deep and shallow traits in order to get a more descriminative face model. Also, we widely evaluated the proposed SILD+WCCN/LR approach versus the state-of-the-art approaches utilizing two challenging facial kinship databases namely KinFaceW-II and TSKinFace.

In chapter 5, we introduced a new local feature descriptor for extracting features from facial images. Our descriptor is based on the local statistics traits of the facial image and the original BSIF operator. Besides, we proposed a novel approach called SIWEDA for face and kinship verification based on the classical SIEDA approach. Furthermore, to lighten the class intra-variability, we proposed two variants SIEDA+WCCN and SIWEDA+WCCN by integrating WCCN in SIEDA and SIWEDA, respectively. We extensively evaluated our approach against the stateof-the-art methods using five challenging face and kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, YTF (video still-frames) and LFW databases.

In chapter 6, we tackled for the first time the kinship verification problem as a cross-view matching problem because every kin relation is typically viewpoint changes from two face images belonging to two different persons. Moreover, we proposed a robust automated facial verification framework suitable for kinship verification, from face images captured in unconstrained environments. The face data is represented as a high order tensor based on the combination of different local features in order to provide a more powerful face model. Furthermore, we proposed a novel method for multi-linear dimensionality reduction and classification, called Tensor Cross-view Quadratic Discriminant Analysis (TXQDA), which preserved the data structure, enlarged the margin between samples, helped lighten the small sample size problem and reduced the computational cost. Also, we extensively evaluate the proposed TXQDA method against the state-of-the-art methods using five challenging kinship databases namely Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II and FIW.

The performance of the TXQDA method was compared to that of subspace metric learning methods counterparts such as SILD, NRML, XQDA, and MSIDA. The proposed method has been validated by several types of analyzes using five kinship databases: Cornell KinFace, UB KinFace, TSkinFace, KinFaceW-II and FIW. The results of the experimental study show that TXQDA is a robust and effective verification method for tensor cross-view matching because it surpasses all the traditional metric learning methods.

7.2 Perspectives

This section presents some ideas for future research in order to design solutions for kinship verification task, as well as to collect new databases in which the proposed methods can be applied. Furthermore, many problems of facial recognition by multidimensional analysis have been addressed in this thesis, the field of multilinear subspaces analysis still presents many unresolved problems which have yet to be taken into account. Therefore, we tried to conclude our perspectives into two big ideas including data collections and tensor subspace analysis for kinship verification.

7.2.1 kinship verification and the lack of big collected data

- Kinship verification is lacking of largest family facial data. Furthermore, the proposed algorithms for kinship verification still stacked in small databases which affect the training performance of models. So, we planned to gather and collect a new database to verify bi-subject matching, tri-subject matching and multi-subject matching (family matching).

- Kinship verification has yet to show an advantage using other modalities than face (i.e. voices, videos and gaits). Also, collecting videos, voices (extracted from videos

for example) and gaits images are one of the most important priorities of our next works.

- The existing kinship verification databases includes only one face for parent and one face for child. This is a very limited case for verifying kinship. In addition to the obstacle generally faced the face verification in unconstrained environments (i.e. facial images captured under uncontrolled environments without any restrictions in terms of pose, lighting, background, expression, and partial occlusion), kinship verification inserts another layer of obstacles which is far from being easy. So, collecting a new database that includes different facial images of parents and their children can extremely help automatic kinship verification algorithms to better understand facial kinship cues.

7.2.2 Development of tensor subspaces analysis metric learning

- Evaluating of a new effective metric learning methods based on multilinear subspaces analysis (tensor-based subspace analysis), which further enlarge the optimisation criterion of our TXQDA method.

- Combining features has become one of the most effective strategies to improve the kinship verification performance. The integration in chapter 6 of TXQDA with the local texture descriptors MSLPQ and MSBSIF showed promising results for facial kinship verification. It would therefore be interesting to study the integration of the feature vectors projected by TXQDA with our StatBIF local statistical features descriptor (proposed in Chapter 5) in order to extract the most discriminating information for better classification.

- Lastly, in multilinear subspaces analysis, there are still many unresolved issues, such as optimal initialization, optimal projection order, and optimal stopping criterion. This thesis has made some attempts to solve some of these problems in the proposed methods TXQDA. However, it will be beneficial if further research leads to a deeper understanding of these issues.

Bibliography

[1] Adini, Y., Moses, Y., Ullman, S.: Face recognition: The problem of compensating for changes in illumination direction. IEEE Transactions on pattern analysis and machine intelligence **19**(7), 721–732 (1997)

[2] Aliradi, R., Belkhir, A., Ouamane, A., Elmaghraby, A.S.: Dieda: discriminative information based on exponential discriminant analysis combined with local features representation for face and kinship verification. Multimedia Tools and Applications pp. 1–18 (2018)

[3] Alirezazadeh, P., Fathi, A., Abdali-Mohammadi, F.: A genetic algorithm-based feature selection for kinship verification. IEEE Signal Processing Letters **22**(12), 2459–2463 (2015). DOI 10.1109/LSP.2015.2490805

[4] Almuashi, M., Hashim, S.Z.M., Mohamad, D., Alkawaz, M.H., Ali, A.: Automated kinship verification and identification through human facial images: a survey. Multimedia Tools and Applications **76**(1), 265–307 (2017)

[5] Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., Raymond, M.: Cross-cultural perceptions of facial resemblance between kin. Journal of Vision **9**(6), 23 (2009). DOI 10.1167/9.6.23. URL `+http://dx.doi.org/10.1167/9.6.23`

[6] Arashloo, S.R., Kittler, J.: Efficient processing of mrfs for unconstrained-pose face recognition. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8 (2013). DOI 10.1109/BTAS.2013.6712721

[7] Arashloo, S.R., Kittler, J.: 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), 2100–2109 (2014). DOI 10.1109/TIFS.2014.2359587

[8] Ayesha, S., Hanif, M.K., Talib, R.: Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion **59**, 44 – 58 (2020). DOI https://doi.org/10.1016/j.inffus.2020.01.005. URL `http://www.sciencedirect.com/science/article/pii/S156625351930377X`

[9] Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: 2013 IEEE International Conference on Computer Vision, pp. 1960–1967 (2013). DOI 10.1109/ICCV.2013.246

[10] Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: 2013 IEEE International Conference on Computer Vision, pp. 1960–1967 (2013). DOI 10.1109/ICCV.2013.246

[11] Bekhouche, S., Ouafi, A., Dornaika, F., Taleb-Ahmed, A., Hadid, A.: Pyramid multilevel features for facial demographic estimation. Expert Systems with Applications **80**, 297 – 310 (2017). DOI https://doi.org/10.1016/j.eswa.2017.03.030. URL `http://www.sciencedirect.com/science/article/pii/S0957417417301793`

[12] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 711–720 (1997). DOI 10.1109/34.598228

[13] Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence **35**(8), 1798–1828 (2013)

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

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

[16] Biran, O., Cotton, C.: Explanation and justification in machine learning: A survey. In: IJCAI-17 workshop on explainable AI (XAI), vol. 8, pp. 8–13 (2017)

[17] Bleicher, P.: Biometrics comes of age: Despite accuracy and security concerns, biometrics are gaining in popularity. Applied Clinical Trials **14**(12), 18–20 (2005)

[18] Bleyer, M., Chambon, S., Poppe, U., Gelautz, M.: Evaluation of different methods for using colour information in global stereo matching approaches. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B3a, pp. 415–420. Vol. XXXVII, Part B3a, Beijing (2008). URL `http://publik.tuwien.ac.at/files/PubDat_169068.pdf`. Vortrag: ISPRS Congress Beijing 2008, Beijing - China; 2008-07-03 – 2008-07-11

[19] Bottino, A., Vieira, T.F., Ul Islam, I.: Geometric and textural cues for automatic kinship verification. International Journal of Pattern Recognition and Artificial Intelligence **29**(03), 1556,001 (2015)

[20] Bottinok, A., Islam, I.U., Vieira, T.F.: A multi-perspective holistic approach to kinship verification in the wild. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 2, pp. 1–6. IEEE (2015)

[21] Carmen, S.: Nist report to the united states congress. summary of nist tandards for biometric accuracy, tamper resistance and interoperabiity [r/ol] (2001)

[22] Chakrabarti, A., Rajagopalan, A.N., Chellappa, R.: Super-resolution of face images using kernel pca-based prior. IEEE Transactions on Multimedia **9**(4), 888–892 (2007). DOI 10.1109/TMM.2007.893346

[23] Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST) **2**(3), 27 (2011). URL `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`

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

[25] Chen, Y.Y., Hsu, W.H., Liao, H.Y.M.: Discovering informative social subgraphs and predicting pairwise relationships from group photos. In: Proceedings of the 20th ACM international conference on Multimedia, pp. 669–678 (2012)

[26] Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Color face recognition for degraded face images. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**(5), 1217–1230 (2009). DOI 10.1109/TSMCB.2009.2014245

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

[28] Dal Martello, M.F., Maloney, L.T.: Where are kin recognition signals in the human face? Journal of Vision **6**(12), 2 (2006). DOI 10.1167/6.12.2. URL `+http://dx.doi.org/10.1167/6.12.2`

[29] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886–893 vol. 1 (2005). DOI 10.1109/CVPR.2005.177

[30] Dandekar, A.R., Nimbarte, M.S.: A survey: Verification of family relationship from parents and child facial images. In: 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, pp. 1–6 (2014)

[31] Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: Proceedings of the 24th International Conference on Machine Learning, ICML '07, pp. 209–216. ACM, New York, NY, USA (2007). DOI 10.1145/1273496.1273523. URL http://doi.acm.org/10.1145/1273496.1273523

[32] Dawson, M., Zisserman, A., Nellåker, C.: From same photo: Cheating on visual kinship challenges. In: Asian Conference on Computer Vision, pp. 654–668. Springer (2018)

[33] DeBruine, L.M., Smith, F.G., Jones, B.C., Roberts, S.C., Petrie, M., Spector, T.D.: Kin recognition signals in adult faces. Vision Research **49**(1), 38 – 43 (2009). DOI https://doi.org/10.1016/j.visres.2008.09.025. URL http://www.sciencedirect.com/science/article/pii/S0042698908004707

[34] Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech, and Language Processing **19**(4), 788–798 (2011). DOI 10.1109/TASL.2010.2064307

[35] Dehghan, A., Ortiz, E.G., Villegas, R., Shah, M.: Who do i look like? determining parent-offspring resemblance via gated autoencoders. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1757–1764 (2014). DOI 10.1109/CVPR.2014.227

[36] Deng, J., Zhou, Y., Zafeiriou, S.: Marginal loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 60–68 (2017)

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

[38] Dibeklioglu, H., Ali Salah, A., Gevers, T.: Like father, like son: Facial expression dynamics for kinship verification. In: Proceedings of the ieee international conference on computer vision, pp. 1497–1504 (2013)

[39] Dornaika, F., Arganda-Carreras, I., Serradilla, O.: Transfer learning and feature fusion for kinship verification. Neural Computing and Applications (2019). DOI 10.1007/s00521-019-04201-0. URL https://doi.org/10.1007/s00521-019-04201-0

[40] Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.: Recognizing faces with pca and ica. Computer Vision and Image Understanding **91**(1), 115 – 137 (2003). DOI https://doi.org/10.1016/S1077-3142(03)00077-8. URL `http://www.sciencedirect.com/science/article/pii/S1077314203000778`. Special Issue on Face Recognition

[41] Duan, Y., Lu, J., Feng, J., Zhou, J.: Learning rotation-invariant local binary descriptor. IEEE Transactions on Image Processing **26**(8), 3636–3651 (2017). DOI 10.1109/TIP.2017.2704661

[42] Duan, Y., Lu, J., Feng, J., Zhou, J.: Context-aware local binary feature learning for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(5), 1139–1153 (2018). DOI 10.1109/TPAMI.2017.2710183

[43] Duan, Y., Lu, J., Wang, Z., Feng, J., Zhou, J.: Learning deep binary descriptor with multi-quantization. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4857–4866 (2017). DOI 10.1109/CVPR.2017.516

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

[45] Fu, Y., Cao, L., Guo, G., Huang, T.S.: Multiple feature fusion by subspace learning. In: Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 127–134 (2008)

[46] Gallagher, A.C., Chen, T.: Understanding images of groups of people. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 256–263. IEEE (2009)

[47] Goldberger, J., Hinton, G.E., Roweis, S.T., Salakhutdinov, R.R.: Neighbourhood components analysis. In: L.K. Saul, Y. Weiss, L. Bottou (eds.) Advances in Neural Information Processing Systems 17, pp. 513–520. MIT Press (2005). URL `http://papers.nips.cc/paper/2566-neighbourhood-components-analysis.pdf`

[48] Gopalan, R., Taheri, S., Turaga, P., Chellappa, R.: A blur-robust descriptor with applications to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(6), 1220–1226 (2012). DOI 10.1109/TPAMI.2012.15

[49] Goyal, A., Meenpal, T.: Eccentricity based kinship verification from facial images in the wild. Pattern Analysis and Applications pp. 1–26 (2020)

[50] Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: D. Forsyth, P. Torr, A. Zisserman (eds.) Computer Vision – ECCV 2008, pp. 262–275. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)

[51] Guehairia, O., Ouamane, A., Dornaika, F., Taleb-Ahmed, A.: Feature fusion via deep random forest for facial age estimation. Neural Networks **130**, 238–252 (2020)

[52] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM computing surveys (CSUR) **51**(5), 1–42 (2018)

[53] Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? metric learning approaches for face identification. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 498–505 (2009). DOI 10.1109/ICCV.2009.5459197

[54] Guo, G., Wang, X.: Kinship measurement on salient facial features. IEEE Transactions on Instrumentation and Measurement **61**(8), 2322–2325 (2012)

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

[56] Harrell Jr, F.E.: Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer (2015)

[57] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA (2001)

[58] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **9**(4), e1312 (2019)

[59] Hu, J., Lu, J., Tan, Y.: Discriminative deep metric learning for face verification in the wild. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1875–1882 (2014). DOI 10.1109/CVPR.2014.242

[60] Hu, J., Lu, J., Tan, Y.: Sharable and individual multi-view metric learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(9), 2281–2288 (2018). DOI 10.1109/TPAMI.2017.2749576

[61] Hu, J., Lu, J., Tan, Y., Yuan, J., Zhou, J.: Local large-margin multi-metric learning for face and kinship verification. IEEE Transactions on Circuits and Systems for Video Technology **28**(8), 1875–1891 (2018). DOI 10.1109/TCSVT.2017.2691801

[62] Hu, J., Lu, J., Tan, Y.P.: Discriminative deep metric learning for face verification in the wild. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14, pp. 1875–1882. IEEE Computer Society,

Washington, DC, USA (2014). DOI 10.1109/CVPR.2014.242. URL `http://dx.doi.org/10.1109/CVPR.2014.242`

[63] Hu, J., Lu, J., Yuan, J., Tan, Y.P.: Large margin multi-metric learning for face and kinship verification in the wild. pp. 252–267 (2014). DOI 10.1007/978-3-319-16811-1_17

[64] Huajie, J., Lichun, W., Yanfeng, S., Yongli, H.: Color face recognition based on color space normalization and quaternion matrix representation. In: 2012 Fourth International Conference on Digital Home, pp. 133–137 (2012). DOI 10.1109/ICDH.2012.19

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

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

[67] Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: 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`

[68] Iman, R.L., Conover, W.J.: A distribution-free approach to inducing rank correlation among input variables. Communications in Statistics - Simulation and Computation **11**(3), 311–334 (1982). DOI 10.1080/03610918208812265. URL `https://doi.org/10.1080/03610918208812265`

[69] Jain, A.K., Bolle, R., Pankanti, S.: Personal identification in networked society. Biometrics (1999)

[70] Jain, A.K., Flynn, P., Ross, A.A.: Handbook of biometrics. Springer Science & Business Media (2007)

[71] Jain, A.K., Ross, A.: Multibiometric systems. Communications of the ACM **47**(1), 34–40 (2004)

[72] Jain, A.K., Ross, A.A., Nandakumar, K.: Introduction to biometrics. Springer Science & Business Media (2011)

[73] Jiang, J., Hu, R., Wang, Z., Han, Z.: Noise robust face hallucination via locality-constrained representation. IEEE Transactions on Multimedia **16**(5), 1268–1281 (2014)

[74] Kaminski, G., Dridi, S., Graff, C., Gentaz, E.: Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. Proceedings of the Royal Society of London B: Biological Sciences **276**(1670), 3193–3200 (2009). DOI 10.1098/rspb.2009.0677. URL http://rspb.royalsocietypublishing.org/content/276/1670/3193

[75] Kaminski, G., Ravary, F., Graff, C., Gentaz, E.: Firstborns' disadvantage in kinship detection. Psychological Science **21**(12), 1746–1750 (2010). DOI 10.1177/0956797610388045. URL http://dx.doi.org/10.1177/0956797610388045. PMID: 21051523

[76] Kan, M., Xu, D., Shan, S., Li, W., Chen, X.: Learning prototype hyperplanes for face verification in the wild. IEEE Transactions on Image Processing **22**, 3310–3316 (2013)

[77] Kannala, J., Rahtu, E.: Bsif: Binarized statistical image features. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pp. 1363–1366 (2012)

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

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

[80] Kohli, N., Yadav, D., Vatsa, M., Singh, R., Noore, A.: Supervised mixed norm autoencoder for kinship verification in unconstrained videos. IEEE Transactions on Image Processing **28**(3), 1329–1341 (2019)

[81] Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2288–2295 (2012). DOI 10.1109/CVPR.2012.6247939

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

[83] Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A., Hadid, A.: Learning multi-view deep and shallow features through new discriminative subspace for bi-subject and tri-subject kinship verification. Applied Intelligence **49**(11), 3894–3908 (2019). DOI 10.1007/s10489-019-01489-2. URL `https://doi.org/10.1007/s10489-019-01489-2`

[84] Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A., Hadid, A.: Tensor cross-view quadratic discriminant analysis for kinship verification in the wild. Neurocomputing (2019). DOI https://doi.org/10.1016/j.neucom.2019.10.055. URL `http://www.sciencedirect.com/science/article/pii/S0925231219314353`

[85] Laiadi, O., Ouamane, A., Boutellaa, E., Benakcha, A., Taleb-Ahmed, A., Hadid, A.: Kinship verification from face images in discriminative subspaces of color components. Multimedia Tools and Applications (2018). DOI 10.1007/s11042-018-7027-9. URL `https://doi.org/10.1007/s11042-018-7027-9`

[86] Lan, R., Zhou, Y.: Quaternion-michelson descriptor for color image classification. IEEE Transactions on Image Processing **25**(11), 5281–5292 (2016)

[87] Lan, R., Zhou, Y., Tang, Y.Y.: Quaternionic weber local descriptor of color images. IEEE Transactions on Circuits and Systems for Video Technology **27**(2), 261–274 (2017)

[88] Lee, T.W.: Independent component analysis. In: Independent component analysis, pp. 27–66. Springer (1998)

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

[90] Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic matching for pose variant face verification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3499–3506 (2013). DOI 10.1109/CVPR.2013.449

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

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

[93] Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person Re-identification by Local Maximal Occurrence Representation and Metric Learning. ArXiv e-prints (2014)

[94] Liao, S., Yi, D., Lei, Z., Qin, R., Li, S.Z.: Heterogeneous face recognition from local structures of normalized appearance. In: M. Tistarelli, M.S. Nixon (eds.) Advances in Biometrics, pp. 209–218. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)

[95] Liu, C.: Discriminant analysis and similarity measure. Pattern Recognition **47**(1), 359 – 367 (2014). DOI https://doi.org/10.1016/j.patcog.2013.06.023. URL `http://www.sciencedirect.com/science/article/pii/S0031320313002756`

[96] Liu, Q., Puthenputhussery, A., Liu, C.: Inheritable fisher vector feature for kinship verification. In: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–6 (2015). DOI 10.1109/BTAS.2015.7358768

[97] Liu, Z., Liu, C.: Fusion of the complementary discrete cosine features in the yiq color space for face recognition. Computer Vision and Image Understanding **111**(3), 249 – 262 (2008). DOI https://doi.org/10.1016/j.cviu.2007.12.002. URL `http://www.sciencedirect.com/science/article/pii/S1077314207001683`

[98] Lopez, M.B., Boutellaa, E., Hadid, A.: Comments on the "kinship face in the wild" data sets. IEEE Trans. Pattern Anal. Mach. Intell. **38**(11), 2342–2344 (2016). DOI 10.1109/TPAMI.2016.2522416. URL `https://doi.org/10.1109/TPAMI.2016.2522416`

[99] Lu, G.F., Wang, Y., Zou, J., Wang, Z.: Matrix exponential based discriminant locality preserving projections for feature extraction. Neural Networks **97**, 127 – 136 (2018). DOI https://doi.org/10.1016/j.neunet.2017.09.014. URL `http://www.sciencedirect.com/science/article/pii/S0893608017302289`

[100] Lu, J., Hu, J., Liong, V.E., Zhou, X., Bottino, A., Islam, I.U., Vieira, T.F., Qin, X., Tan, X., Chen, S., Mahpod, S., Keller, Y., Zheng, L., Idrissi, K., Garcia, C., Duffner, S., Baskurt, A., Castrillón-Santana, M., Lorenzo-Navarro, J.: The fg 2015 kinship verification in the wild evaluation. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1, pp. 1–7 (2015). DOI 10.1109/FG.2015.7163159

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

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

[103] Lu, J., Zhou, X., Tan, Y.P., Shang, Y., Zhou, J.: Neighborhood repulsed metric learning for kinship verification. IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 331–345 (2014). DOI 10.1109/TPAMI.2013.134. URL http://dx.doi.org/10.1109/TPAMI.2013.134

[104] Ma, J., Zhao, Y., Ahalt, S.: Osu svm classifier matlab toolbox (ver 3.00). Pulsed Neural Networks (2002). URL http://svm.sourceforge.net/docs/3.00/api/

[105] Mahpod, S., Keller, Y.: Kinship verification using multiview hybrid distance learning. Computer Vision and Image Understanding **167**, 28 – 36 (2018). DOI https://doi.org/10.1016/j.cviu.2017.12.003. URL http://www.sciencedirect.com/science/article/pii/S107731421730228X

[106] Mao, Q., Rao, Q., Yu, Y., Dong, M.: Hierarchical bayesian theme models for multipose facial expression recognition. IEEE Transactions on Multimedia **19**(4), 861–873 (2017). DOI 10.1109/TMM.2016.2629282

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

[108] Mehrotra, R., Namuduri, K.R., Ranganathan, N.: Gabor filter-based edge detection. Pattern recognition **25**(12), 1479–1494 (1992)

[109] Meina Kan Shiguang Shan, D.X., Chen, X.: Side-information based linear discriminant analysis for face recognition. In: Proc. BMVC, pp. 125.1–125.0 (2011). Http://dx.doi.org/10.5244/C.25.125

[110] Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence **267**, 1 – 38 (2019). DOI https://doi.org/10.1016/j.artint.2018.07.007. URL http://www.sciencedirect.com/science/article/pii/S0004370218305988

[111] Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognition **33**(11), 1771 – 1782 (2000). DOI https://doi.org/10.1016/S0031-3203(99)00179-X. URL http://www.sciencedirect.com/science/article/pii/S003132039900179X

[112] Molnar, C.: Interpretable machine learning: a guide for making black box models explainable. leanpub (2018)

[113] Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S.: Agedb: the first manually collected, in-the-wild age database. In: Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 51–59 (2017)

[114] Moujahid, A., Dornaika, F.: A pyramid multi-level face descriptor: application to kinship verification. Multimedia Tools and Applications (2018). DOI 10.1007/s11042-018-6517-0. URL https://doi.org/10.1007/s11042-018-6517-0

[115] Méndez-Vázquez, H., Martínez-Díaz, Y., Chai, Z.: Volume structured ordinal features with background similarity measure for video face recognition. In: 2013 International Conference on Biometrics (ICB), pp. 1–6 (2013). DOI 10.1109/ICB.2013.6612990

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

[117] Nosaka, R., Ohkawa, Y., Fukui, K.: Feature Extraction Based on Co-occurrence of Adjacent Local Binary Patterns, pp. 82–91. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). DOI 10.1007/978-3-642-25346-1_8. URL http://dx.doi.org/10.1007/978-3-642-25346-1_8

[118] Nowak, E., Jurie, F.: Learning visual similarity measures for comparing never seen objects. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007). DOI 10.1109/CVPR.2007.382969

[119] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7), 971–987 (2002). DOI 10.1109/TPAMI.2002.1017623

[120] Ojansivu, V., Heikkilä, J.: Blur Insensitive Texture Classification Using Local Phase Quantization, pp. 236–243. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). DOI 10.1007/978-3-540-69905-7_27. URL http://dx.doi.org/10.1007/978-3-540-69905-7_27

[121] Ouamane, A., Bengherabi, M., Hadid, A., Cheriet, M.: Side-information based exponential discriminant analysis for face verification in the wild. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 02, pp. 1–6 (2015). DOI 10.1109/FG.2015.7284837

[122] Ouamane, A., Boutellaa, E., Bengherabi, M., Taleb-Ahmed, A., Hadid, A.: A novel statistical and multiscale local binary feature for 2d and 3d face verification. Computers & Electrical Engineering **62**, 68–80 (2017)

[123] Ouamane, A., Chouchane, A., Boutellaa, E., Belahcene, M., Bourennane, S., Hadid, A.: Efficient tensor-based 2d+3d face verification. IEEE Transactions on Information Forensics and Security **PP**(99), 1–1 (2017). DOI 10.1109/TIFS.2017.2718490

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

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

[126] Owusu, E., Zhan, Y., Mao, Q.R.: An svm-adaboost facial expression recognition system. Applied Intelligence **40**(3), 536–545 (2014). DOI 10.1007/s10489-013-0478-9. URL `https://doi.org/10.1007/s10489-013-0478-9`

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

[128] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference (2015)

[129] Peleg, G., Katzir, G., Peleg, O., Kamara, M., Brodsky, L., Hel-Or, H., Keren, D., Nevo, E.: Hereditary family signature of facial expression. Proceedings of the National Academy of Sciences **103**(43), 15,921–15,926 (2006)

[130] Pinto, N., DiCarlo, J.J., Cox, D.D.: How far can you get with a modern face recognition test set using only simple features? In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2591–2598 (2009). DOI 10.1109/CVPR.2009.5206605

[131] Qin, X., Liu, D., Wang, D.: Heterogeneous similarity learning for more practical kinship verification. Neural Processing Letters **47**(3), 1253–1269 (2018). DOI 10.1007/s11063-017-9694-3. URL `https://doi.org/10.1007/s11063-017-9694-3`

[132] Qin, X., Liu, D., Wang, D.: A literature survey on kinship verification through facial images. Neurocomputing **377**, 213 – 224 (2020). DOI https://doi.org/10.1016/j.neucom.2019.09.089. URL `http://www.sciencedirect.com/science/article/pii/S0925231219314328`

[133] Qin, X., Tan, X., Chen, S.: Tri-subject kinship verification: Understanding the core of a family. IEEE Transactions on Multimedia **17**(10), 1855–1867 (2015). DOI 10.1109/TMM.2015.2461462

[134] Qin, X., Tan, X., Chen, S.: Mixed bi-subject kinship verification via multi-view multi-task learning. Neurocomputing **214**, 350 – 357 (2016). DOI https://doi.org/10.1016/j.neucom.2016.06.027. URL `http://www.sciencedirect.com/science/article/pii/S0925231216306658`

[135] Ramanathan, N., Chellappa, R.: Modeling age progression in young faces. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, pp. 387–394. IEEE (2006)

[136] Raudys, S.J., Jain, A.K.: Small sample size effects in statistical pattern recognition: recommendations for practitioners. IEEE Transactions on Pattern Analysis and Machine Intelligence **13**(3), 252–264 (1991). DOI 10.1109/34.75512

[137] Rettkowski, J., Boutros, A., Göhringer, D.: Hw/sw co-design of the hog algorithm on a xilinx zynq soc. Journal of Parallel and Distributed Computing **109**, 50–62 (2017)

[138] Robinson, J.P., Shao, M., Wu, Y., Liu, H., Gillis, T., Fu, Y.: Visual kinship recognition of families in the wild. In: IEEE Transactions on pattern analysis and machine intelligence (2018)

[139] Sahoo, S.K., Choubisa, T., Prasanna, S.M.: Multimodal biometric person authentication: A review. IETE Technical Review **29**(1), 54–75 (2012)

[140] Sanderson, C., Lovell, B.C.: Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference, pp. 199–208. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). DOI 10.1007/978-3-642-01793-3_21. URL `http://dx.doi.org/10.1007/978-3-642-01793-3_21`

[141] Seo, H.J., Milanfar, P.: Face verification using the lark representation. IEEE Transactions on Information Forensics and Security **6**(4), 1275–1286 (2011)

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

[143] Song, H.O., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4004–4012 (2016). DOI 10.1109/CVPR.2016.434

[144] Sturm, R.A., Box, N.F., Ramsay, M.: Human pigmentation genetics: the difference is only skin deep. BioEssays **20**(9), 712–721 (1998). DOI 10.1002/(SICI)1521-1878(199809)20:9<712::AID-BIES4>3.0.CO;2-I. URL `http://dx.doi.org/10.1002/(SICI)1521-1878(199809)20:9<712::AID-BIES4>3.0.CO;2-I`

[145] Swets, D.L., Weng, J.J.: Using discriminant eigenfeatures for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence **18**(8), 831–836 (1996). DOI 10.1109/34.531802

[146] Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014). DOI 10.1109/CVPR.2014.220

[147] Torres, L., Reutter, J.Y., Lorente, L.: The importance of the color information in face recognition. In: Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), vol. 3, pp. 627–631 vol.3 (1999). DOI 10.1109/ICIP.1999.817191

[148] Touil, D.E., Terki, N., Medouakh, S.: Learning spatially correlation filters based on convolutional features via pso algorithm and two combined color spaces for visual tracking. Applied Intelligence **48**(9), 2837–2846 (2018)

[149] Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of cognitive neuroscience **3**(1), 71–86 (1991)

[150] Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586–591 (1991). DOI 10.1109/CVPR.1991.139758

[151] Vieira, T.F., Bottino, A., Laurentini, A., De Simone, M.: Detecting siblings in image pairs. The Visual Computer **30**(12), 1333–1345 (2014)

[152] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1, pp. I–I. IEEE (2001)

[153] Wang, G., Gallagher, A., Luo, J., Forsyth, D.: Seeing people in social context: Recognizing people and social relationships. In: European conference on computer vision, pp. 169–182. Springer (2010)

[154] Wang, J., Barreto, A., Wang, L., Chen, Y., Rishe, N., Andrian, J., Adjouadi, M.: Multilinear principal component analysis for face recognition with fewer features. Neurocomputing **73**(10), 1550 – 1555 (2010). DOI https://doi.org/10.1016/j.neucom.2009.08.022. URL `http://www.sciencedirect.com/science/article/`

pii/S092523121000113X. Subspace Learning / Selected papers from the European Symposium on Time Series Prediction

[155] Wang, S., Ding, Z., Fu, Y.: Cross-generation kinship verification with sparse discriminative metric. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(11), 2783–2790 (2019)

[156] Wang, S., Robinson, J.P., Fu, Y.: Kinship verification on families in the wild with marginalized denoising metric learning. In: Automatic Face and Gesture Recognition (FG), 2017 12th IEEE International Conference and Workshops on

[157] Wang, S., Yan, H.: Discriminative sampling via deep reinforcement learning for kinship verification. Pattern Recognition Letters **138**, 38 – 43 (2020). DOI https://doi.org/10.1016/j.patrec.2020.06.019. URL http://www.sciencedirect.com/science/article/pii/S0167865520302373

[158] Wei, W., Dai, H., tai Liang, W.: Exponential sparsity preserving projection with applications to image recognition. Pattern Recognition **104**, 107,357 (2020). DOI https://doi.org/10.1016/j.patcog.2020.107357. URL http://www.sciencedirect.com/science/article/pii/S0031320320301606

[159] Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10**, 207–244 (2009). URL http://dl.acm.org/citation.cfm?id=1577069.1577078

[160] Wilson, C., Hicklin, A., Bone, M., Korves, H., Grother, P., Ulery, B., Micheals, R., Zoepfl, M., Otto, S., Watson, C.: Fingerprint vendor technology evaluation 2003: summary of results and analysis report. NIST Technical Report NISTIR **7123** (2004)

[161] Wilson, C.L., Garris, M.D., Watson, C.I.: Matching performance for the us-visit ident system using flat fingerprints. In: National Institute of Standards and Technology Internal Report 7110. Citeseer (2004)

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

[163] Wolf, L., Hassner, T., Taigman, Y.: 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

[164] Wong, J.J., Cho, S.Y.: A face emotion tree structure representation with probabilistic recursive neural network modeling. Neural Computing and Applications **19**(1), 33–54 (2010)

[165] Wu, F., Jing, X.Y., Dong, X., Ge, Q., Wu, S., Liu, Q., Yue, D., Yang, J.Y.: Uncorrelated multi-set feature learning for color face recognition. Pattern Recognition **60**, 630 – 646 (2016). DOI https://doi.org/10.1016/j.patcog.2016.06.010. URL http://www.sciencedirect.com/science/article/pii/S0031320316301261

[166] Wu, G., ting Feng, T., jia Zhang, L., Yang, M.: Inexact implementation using krylov subspace methods for large scale exponential discriminant analysis with applications to high dimensionality reduction problems. Pattern Recognition **66**, 328 – 341 (2017). DOI https://doi.org/10.1016/j.patcog.2016.08.020. URL http://www.sciencedirect.com/science/article/pii/S0031320316302333

[167] Wu, P., Hoi, S.C., Xia, H., Zhao, P., Wang, D., Miao, C.: Online multimodal deep similarity learning with application to image retrieval. In: Proceedings of the 21st ACM International Conference on Multimedia, MM '13, pp. 153–162. ACM, New York, NY, USA (2013). DOI 10.1145/2502081.2502112. URL http://doi.acm.org/10.1145/2502081.2502112

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

[169] Xia, S., Shao, M., Fu, Y.: Kinship verification through transfer learning. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three, IJCAI'11, pp. 2539–2544. AAAI Press (2011). DOI 10.5591/978-1-57735-516-8/IJCAI11-422. URL http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-422

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

[171] Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning, with application to clustering with side-information. In: Proceedings of the 15th International Conference on Neural Information Processing Systems, NIPS'02, pp. 521–528. MIT Press, Cambridge, MA, USA (2002). URL http://dl.acm.org/citation.cfm?id=2968618.2968683

[172] Yan, H.: Kinship verification using neighborhood repulsed correlation metric learning. Image Vision Comput. **60**(C), 91–97 (2017). DOI 10.1016/j.imavis.2016.08.009. URL `https://doi.org/10.1016/j.imavis.2016.08.009`

[173] Yan, H.: Learning discriminative compact binary face descriptor for kinship verification. Pattern Recognition Letters **117**, 146 – 152 (2019). DOI https://doi.org/10.1016/j.patrec.2018.05.027. URL `http://www.sciencedirect.com/science/article/pii/S0167865518302204`

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

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

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

[177] Yan, H., Song, C.: Multi-scale deep relational reasoning for facial kinship verification. Pattern Recognition p. 107541 (2020). DOI https://doi.org/10.1016/j.patcog.2020.107541. URL `http://www.sciencedirect.com/science/article/pii/S0031320320303447`

[178] Yan, H., Wang, S.: Learning part-aware attention networks for kinship verification. Pattern Recognition Letters **128**, 169 – 175 (2019). DOI https://doi.org/10.1016/j.patrec.2019.08.023. URL `http://www.sciencedirect.com/science/article/pii/S0167865519302351`

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

[180] Yang, H., Wang, X.A.: Cascade classifier for face detection. Journal of Algorithms & Computational Technology **10**(3), 187–197 (2016)

[181] Yang, J., Liu, C.: Color image discriminant models and algorithms for face recognition. IEEE Transactions on Neural Networks **19**(12), 2088–2098 (2008). DOI 10.1109/TNN.2008.2003187

[182] Yang, M., Van Gool, L., Zhang, L.: Sparse variation dictionary learning for face recognition with a single training sample per person. In: Proceedings of the IEEE international conference on computer vision, pp. 689–696 (2013)

[183] Yip, A.W., Sinha, P.: Contribution of color to face recognition. Perception **31**(8), 995–1003 (2002). DOI 10.1068/p3376. URL http://dx.doi.org/10.1068/p3376. PMID: 12269592

[184] Yu, H., Chung, C.Y., Wong, K.P., Lee, H.W., Zhang, J.H.: Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition. IEEE Transactions on Power Systems **24**(2), 661–667 (2009). DOI 10.1109/TPWRS.2009.2016589

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

[186] Yuan, S., Mao, X.: Exponential elastic preserving projections for facial expression recognition. Neurocomputing **275**, 711 – 724 (2018). DOI https://doi.org/10.1016/j.neucom.2017.08.067. URL http://www.sciencedirect.com/science/article/pii/S0925231217315126

[187] Zhang, K., Huang, Y., Song, C., Wu, H., Wang, L.: Kinship verification with deep convolutional neural networks. In: BMVC (2015)

[188] Zhang, K., Huang, Y., Song, C., Wu, H., Wang, L., Intelligence, S.M.: Kinship verification with deep convolutional neural networks (2015)

[189] Zhang, L., Duan, Q., Zhang, D., Jia, W., Wang, X.: Advkin: Adversarial convolutional network for kinship verification. IEEE Transactions on Cybernetics pp. 1–14 (2020)

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

[191] Zhang, Z., Chen, Y., Saligrama, V.: Group membership prediction. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 3916–3924 (2015). DOI 10.1109/ICCV.2015.446

[192] Zhao, M., Zhang, Z., Chow, T.W., Li, B.: Soft label based linear discriminant analysis for image recognition and retrieval. Computer Vision and Image Understanding **121**(Supplement C), 86 – 99 (2014). DOI https://doi.org/10.1016/j.

cviu.2014.01.008. URL http://www.sciencedirect.com/science/article/pii/S1077314214000150

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

[194] Zhou, X., Hu, J., Lu, J., Shang, Y., Guan, Y.: Kinship verification from facial images under uncontrolled conditions. In: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, pp. 953–956. ACM, New York, NY, USA (2011). DOI 10.1145/2072298.2071911. URL http://doi.acm.org/10.1145/2072298.2071911

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

[196] Zhou, X., Lu, J., Hu, J., Shang, Y.: Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments. In: Proceedings of the 20th ACM International Conference on Multimedia, MM '12, pp. 725–728. ACM, New York, NY, USA (2012). DOI 10.1145/2393347.2396297. URL http://doi.acm.org/10.1145/2393347.2396297

[197] Zhou, X., Shang, Y., Yan, H., Guo, G.: Ensemble similarity learning for kinship verification from facial images in the wild. Inf. Fusion **32**(PB), 40–48 (2016). DOI 10.1016/j.inffus.2015.08.006. URL https://doi.org/10.1016/j.inffus.2015.08.006

A Scientific productions

A.1 Publications in journals

- Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A., & Hadid, A. (2021). A weighted exponential discriminant analysis through side-information for face and kinship verification using statistical binarized image features. International Journal of Machine Learning and Cybernetics, 12, 171-185. doi: 10.1007/s13042-020-01163-x.

- Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A., & Hadid, A. (2020). Tensor cross-view quadratic discriminant analysis for kinship verification in the wild. Neurocomputing, 377, 286-300. doi: 10.1016/j.neucom.2019.10.055.

- Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A., & Hadid, A. (2019). Learning multi-view deep and shallow features through new discriminative subspace for bi-subject and tri-subject kinship verification. Applied Intelligence, 49(11), 3894-3908. doi: 10.1007/s10489-019-01489-2.

- Laiadi, O., Ouamane, A., Boutellaa, E., Benakcha, A., Taleb-Ahmed, A., & Hadid, A. (2019). Kinship verification from face images in discriminative subspaces of color components. Multimedia Tools and Applications, 78(12), 16465-16487. doi: 10.1007/s11042-018-7027-9.

A.2 Publications in international conferences

- O. Laiadi, A. Ouamane, A. Benakcha, A. Taleb-ahmed and 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) (FG), Buenos Aires, Argentina, 2020, pp. 743-747. doi: 10.1109/FG47880.2020.00118.

- O. Laiadi, A. Ouamane, A. Benakcha, A. Taleb-Ahmed and A. Hadid, "Kinship Verification based Deep and Tensor Features through Extreme Learning Machine," 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 2019, pp. 1-4, doi: 10.1109/FG.2019.8756627.

- Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A., & Hadid, A. Local Phase Quantization Features Extraction in Discriminative Subspace for Kinship

Verification. 2017 5th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2017. (**Best paper award**).

- Oualid Laiadi, Abdelmalik Ouamane, Abdelhamid Benakcha, and Abdelmalik Taleb-Ahmed. 2017. RFIW 2017: LPQ-SIEDA for Large Scale Kinship Verification. In Proceedings of the 2017 Workshop on Recognizing Families In the Wild (RFIW'17). Association for Computing Machinery, New York, NY, USA, 37–39. DOI: 10.1145/3134421.3134426.

B Performance metrics

		Predicted condition			
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP), hit**	**False negative (FN), type II error, miss, underestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
	Negative (N)	**False positive (FP), type I error, false alarm, overestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2\,PPV \times TPR}{PPV + TPR} = \frac{2\,TP}{2\,TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

Figure B.1: The relationship between sensitivity, specificity, and similar terms can be understood using the following table. Consider a group with P positive instances and N negative instances of some condition. The four outcomes can be formulated in a 2×2 contingency table or confusion matrix, as well as derivations of several metrics using the four outcomes.

C Notations and concepts of tensor algebra

This section focuses on the algebra of the multilinear Cross-view Quadratic Discriminant Analysis (TXQDA) method. The variables and mathematical notations that we used in our work are as follows : Lowercase and uppercase symbols (e.g., i, j, F, N and V) indicate scalars; Bold lowercase symbols (e.g., $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$) indicate vectors; italic uppercase symbols (e.g., U, X, Y and W) indicate matrices; bold italic uppercase symbols (e.g., $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$) indicate tensors. A tensor is explained as a multidimensional array [82, 179]. N is considered the order of the tensor and $\boldsymbol{X}$ is called an N^{th}-order tensor. $\mathrm{I_k}$, $1 \leq \mathrm{k} \leq \mathrm{N}$, is the dimension of the k^{th} mode.

The following definitions explain the mathematical tools used to deal with the high order tensors

Definition 1. The inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle \in \Re^{\mathrm{I_1} \times \mathrm{I_2} \times \cdots \times \mathrm{I_N}}$ of two tensors $\boldsymbol{X}$ and $\boldsymbol{Y}$ which have the same order and dimensions is defined by:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{\mathrm{i_1}=1 \ldots \mathrm{i_N}=1}^{\mathrm{I_1} \ldots \mathrm{I_N}} \mathtt{X}_{\mathrm{i_1} \ldots \mathrm{i_N}} \mathtt{Y}_{\mathrm{i_1} \ldots \mathrm{i_N}} \tag{C.1}$$

The Frobenius norm of a tensor $\boldsymbol{X} \in \Re^{\mathrm{I_1} \times \mathrm{I_2} \times \cdots \times \mathrm{I_N}}$ is defined as $\|\boldsymbol{X}\|_{\mathrm{F}} = \sqrt{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}$, and the Euclidean distance between two tensors $\boldsymbol{X}$, $\boldsymbol{Y} \in \Re^{\mathrm{I_1} \times \mathrm{I_2} \times \cdots \times \mathrm{I_N}}$ is defined as $\mathrm{D}(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X} \text{-} \boldsymbol{Y}\|_{\mathrm{F}}$.

Definition 2. The k-mode flattening a tensor $\boldsymbol{B} \in \Re^{\mathrm{I_1} \times \mathrm{I_2} \times \cdots \times \mathrm{I_N}}$ to a matrix $B^{(\mathrm{k})} \in \Re^{\mathrm{I_k} \times \prod_{\mathrm{i} \neq \mathrm{k}} \mathrm{I_i}}$ is defined by:

$$B^{\mathrm{k}} \Leftarrow_{\mathrm{k}} \boldsymbol{B} \tag{C.2}$$

Where

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

The unfolding operation on a 3^{rd}-order tensor is illustrated by Fig. C.1.

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

$$\boldsymbol{Y}_{\mathrm{i_1}, \ldots, \mathrm{i_{k-1}}, \mathrm{i}, \mathrm{i_{k+1}}, \ldots, \mathrm{i_N}} = \sum_{\mathrm{j}=1}^{\mathrm{I_k}} \boldsymbol{X}_{\mathrm{i_1}, \ldots, \mathrm{i_{k-1}}, \mathrm{i}, \mathrm{i_{k+1}}, \ldots, \mathrm{i_N}} G_{\mathrm{i,j}} \tag{C.4}$$

With $\mathrm{j} = 1, \ldots, \mathrm{I_k}$ in which $G_{\mathrm{i,j}}$ indicates the element in the matrix G of coordinates

(i, j). Fig. C.2 illustrates an example of 1-mode vector product of third-order tensor $\boldsymbol{X} \in \Re^{300\times6\times4}$ with matrix $G^{\mathrm{T}} \in \Re^{4\times300}$, resulting tensor $\boldsymbol{X} \times_1 G^{\mathrm{T}} \in \Re^{4\times6\times4}$.

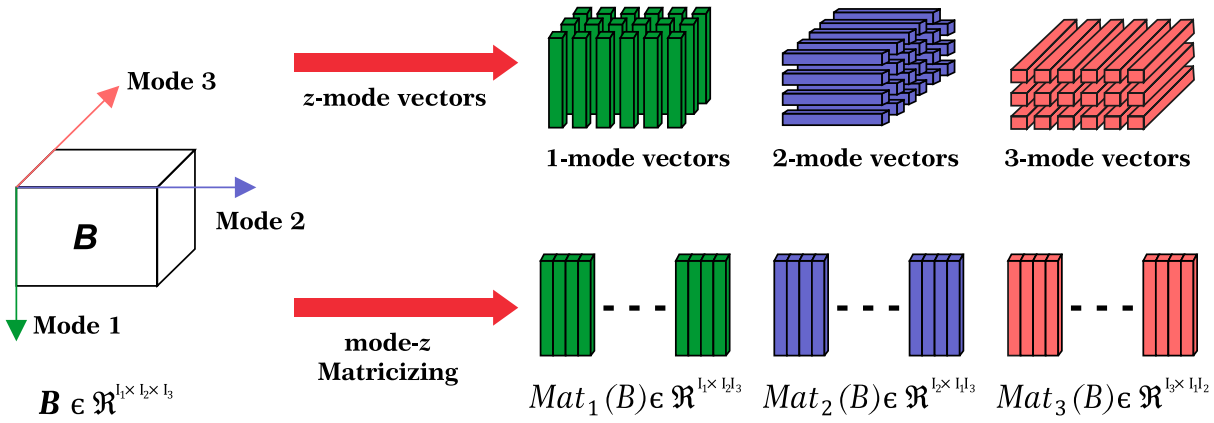


Figure C.1: Example of tensor unfolding.

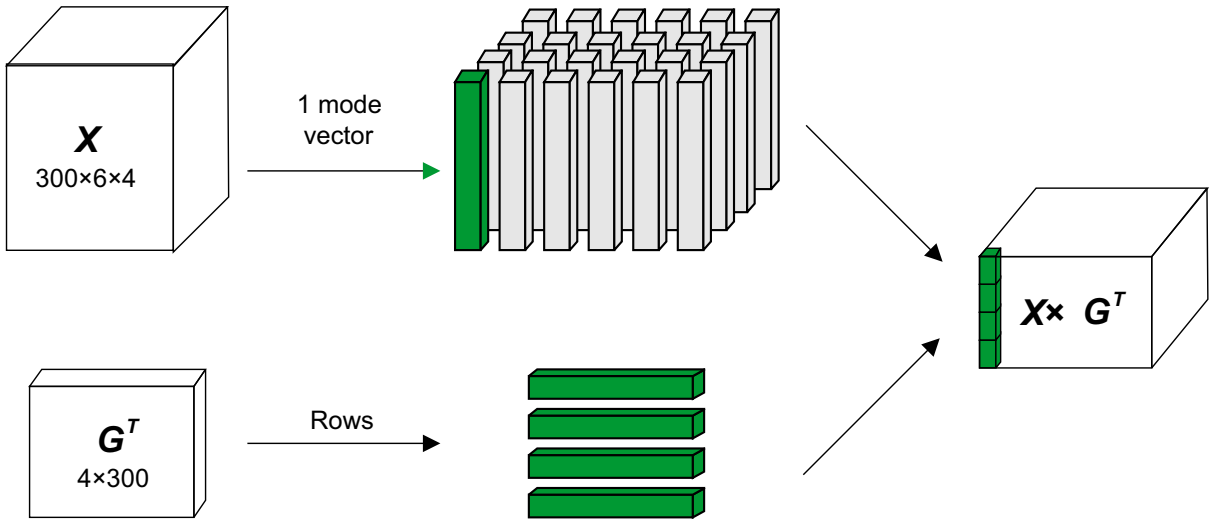


Figure C.2: Visual illustration of 1-mode vector product of third-order tensor $\boldsymbol{X} \in \Re^{300\times6\times4}$ with matrix $G^{\mathrm{T}} \in \Re^{4\times300}$.

D Detailed results of color spaces on the four databases

Tables D.1, D.2, D.3 and D.4 provide the detailed mean accuracy of kinship verification of different descriptors and color spaces on the Cornell KinFace, TSKinFace, KinFaceW-I and KinFaceW-II databases, respectively. The values in the tables are the kinship verification rates (accuracy in %).

Table D.1: The mean accuracy (%) of kinship verification on Cornell database.

Descriptor \ Color space	HSL	HSV	Lab	Luv	RGB	YCbCr	YUV	Gray
MSBSIF	80.45	**81.50**	76.57	76.95	76.19	78.01	77.68	74.48
MSLPQ	79.04	78.71	79.13	77.71	77.64	77.65	78.38	76.94
MSCoALBP	76.83	76.16	73.07	73.72	71.31	76.56	74.42	71.01
Fusion1	80.11	80.48	79.11	78.38	77.60	77.32	77.63	76.25
Fusion2	79.70	79.68	77.28	76.24	75.51	79.06	77.31	73.77
Fusion3	79.34	78.96	77.65	77.98	77.26	78.70	79.07	**77.60**
Fusion4	79.32	79.37	77.65	77.63	77.89	77.98	79.77	75.51

Table D.2: The mean accuracy (%) of kinship verification on TSKinFace database.

Descriptor \ Color space	HSL	HSV	Lab	Luv	RGB	YCbCr	YUV	Gray
MSBSIF	87.54	87.74	86.34	86.78	83.47	87.44	87.32	82.39
MSLPQ	86.58	86.80	86.40	86.21	83.89	86.83	86.16	83.23
MSCoALBP	81.87	81.54	77.02	77.95	78.78	77.86	81.26	76.23
Fusion1	87.76	87.64	86.95	86.96	84.06	87.32	87.25	82.98
Fusion2	87.72	**88.11**	86.48	86.66	84.46	87.52	87.54	82.59
Fusion3	87.02	87.74	86.46	86.68	84.88	86.97	87.66	**83.63**
Fusion4	87.57	88.01	86.88	86.53	84.56	87.30	87.59	83.28

Table D.3: The mean accuracy (%) of kinship verification on KinFaceW-I database.

Descriptor \ Color space	HSL	HSV	Lab	Luv	RGB	YCbCr	YUV	Gray
MSBSIF	78.19	78.98	78.11	78.70	77.55	78.41	78.05	76.50
MSLPQ	77.51	76.94	77.44	77.40	76.75	77.36	77.77	76.30
MSCoALBP	74.44	73.80	73.50	72.50	72.36	72.41	73.04	71.33
Fusion1	78.13	78.85	78.04	78.97	77.42	78.18	78.35	76.91
Fusion2	79.18	79.91	78.88	79.19	78.53	78.21	78.22	77.43
Fusion3	78.91	79.55	78.77	79.07	78.33	78.31	78.71	77.28
Fusion4	78.86	**80.00**	78.98	79.44	78.71	78.11	79.06	**77.59**

Table D.4: The mean accuracy (%) of kinship verification on KinFaceW-II database.

Descriptor \ Color space	HSL	HSV	Lab	Luv	RGB	YCbCr	YUV	Gray
MSBSIF	87.50	86.85	83.85	84.35	78.55	84.45	84.50	77.45
MSLPQ	86.15	85.70	82.85	82.50	78.20	84.55	83.00	77.00
MSCoALBP	79.60	81.00	74.75	74.50	74.05	79.00	77.60	72.25
Fusion1	87.50	87.35	84.00	84.90	79.00	84.50	84.65	77.35
Fusion2	**87.60**	87.40	83.75	84.15	79.30	85.05	84.50	77.85
Fusion3	86.65	86.55	83.25	83.45	79.95	85.10	84.15	**78.05**
Fusion4	87.55	87.30	84.45	84.45	80.15	85.00	84.25	77.95

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES HAUTS-DE-FRANCE

Université Polytechnique HAUTS-DE-FRANCE

UNIVERSITY OF BISKRA

numéro d'ordre: 21/08

Thèse de doctorat

Pour obtenir le grade de Docteur de l'UNIVERSITÉ POLYTECHNIQUE HAUTS-DE-FRANCE, l'INSA HAUTS-DE-FRANCE et l'UNIVERSITÉ MOHAMMED KHIDER DE BISKRA

Résumé de la thèse

Vérification de la parenté entre deux personnes par apprentissage machine

Présentée et soutenue par **Oualid Laiadi**

Le: 13/09/2021, Biskra

Ecole doctorale:

École Doctorale Polytechnique Hauts-de-France (ED PHF)

Directeur de thèse:

Mr. Abdelmalik Taleb-Ahmed	Professeur	Université de Hauts-de-France (France)	Co-directeur
Mr. Abdelhamid Benakcha	Professeur	Université de Biskra (Algeria)	Co-directeur

Jury:

Ms. Malika Mimi	Professeur	Université de Mostaganem (Algérie)	Présidente
Ms. Nacéra Benamrane	Professeur	Université d'Oran (Algérie)	Rapporteurs
Mr. Kamel Eddine Melkemi	Professeur	Université de Batna (Algérie)	Rapporteurs
Mr. Amir Nakib	MC HDR	Université de Créteil (France)	Rapporteurs
Ms. Salima Ouadfel	MC HDR	Université de Constantine (Algérie)	Examinateurs
Mr. Abdeljalil Ouahabi	Professeur	Université de Tours (France)	Examinateurs
Ms. Atika Rivenq	Professeur	Université de Hauts-de-France (France)	Invité
Mr. Abdenour Hadid	Professeur	Université d'Oulu (Finland)	Invité

Table des matières

Table des figures

1 Résumé

1.1 Contexte et motivation

La capacité de déterminer l'identité des personnes et d'accorder des traits personnels (par exemple, nom, âge, nationalité, etc.) à une personne a été très intrinsèque à la structure de notre société. En général, les humains ont utilisé des caractéristiques d'apparence telles que la voix, le visage et la démarche ainsi que d'autres informations contextuelles (par exemple, les vêtements et l'emplacement) pour s'identifier. L'ensemble des traits associés à un individu décrit sa propre identité personnelle. Au début de la civilisation, les gens vivaient dans de petites communautés limitées où les personnes pouvaient facilement s'identifier. En outre, une grande explosion de l'expansion de la population accompagnée d'une mobilité accrue dans la société moderne qui a nécessité le développement de systèmes avancés de gestion automatique des identités capables d'enregistrer, de préserver et d'effacer efficacement les identités privées des peuples.

Un système biométrique facial est subdivisé en deux phases, la phase de apprentissage (apprentissage hors ligne) et la phase de test (classification / vérification en ligne). L'étape de apprentissage ne sera effectuée qu'une seule fois au cours de laquelle l'enrôlement des images faciales des différents individus est utilisé afin d'extraire et de décrire la signature biométrique de chaque individu. Lors de la phase de test, les nouvelles données sont comparées aux données de apprentissage automatiquement apprises lors de la phase de apprentissage afin de prendre la décision d'accepter ou de rejeter le candidat. Les étapes effectuées dans ces deux phases dans un système de reconnaissance faciale sont subdivisées en trois modules principaux [14]: détection de visage, extraction des traits et reconnaissance (classification). Le schéma général du système de reconnaissance faciale est illustré dans Fig. 1.1. Un schéma détaillé du système de reconnaissance faciale est illustré dans Fig. 1.2.

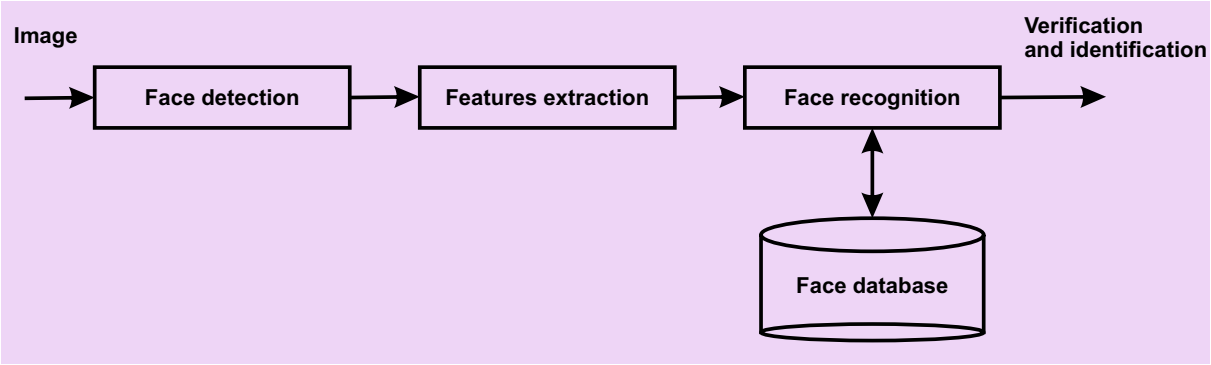


Figure 1.1: Structure du système de reconnaissance faciale.

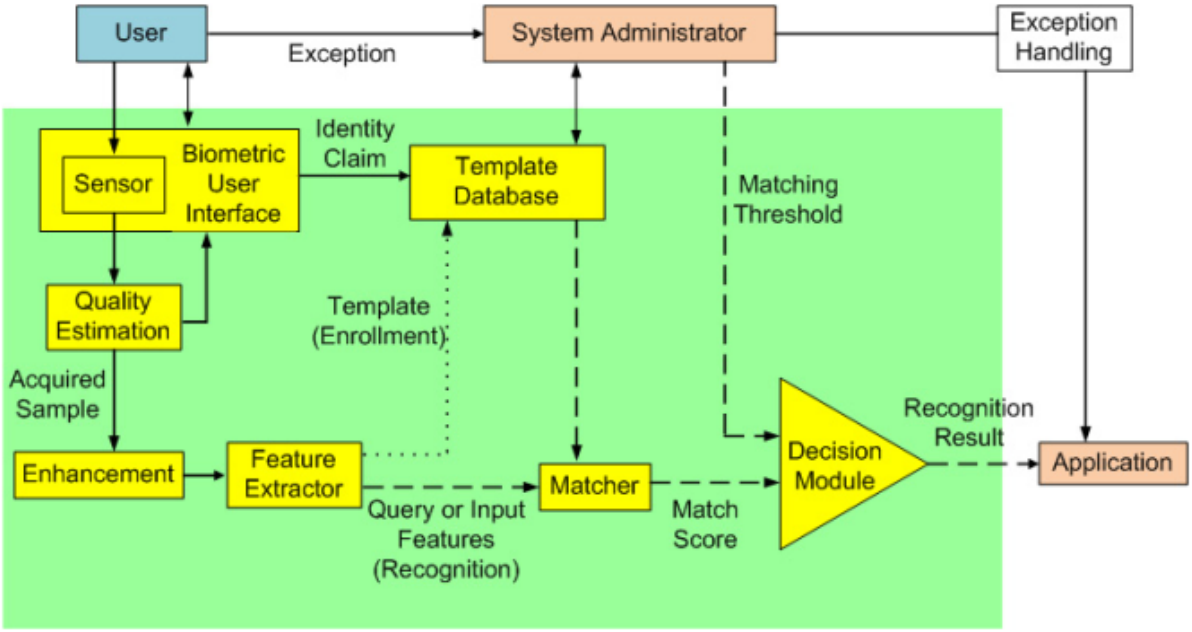


Figure 1.2: Éléments de base d'un système biométrique générique [16].

Trois étapes essentielles sont utilisées pour développer un système de reconnaissance faciale robuste: (i) la détection du visage, (ii) l'extraction de caractéristiques et (iii) la reconnaissance faciale (illustrée à la figure 1.1). L'étape de détection de visage est utilisée pour détecter et déterminer l'image faciale humaine acquise par le système. L'étape d'extraction de caractéristiques est utilisée pour extraire les données de caractéristiques discriminantes pour chaque visage humain déterminé dans la première étape. Enfin, l'étape de reconnaissance faciale comprend les caractéristiques extraites du visage humain qui le comparent à des bases de données faciales complètes pour déterminer l'identité faciale humaine.

- **Détection facial:** Le système de reconnaissance faciale commence d'abord par le centre des soins du visage humains dans une image d'entrée. Le but de cette étape est de définir si l'image cible comprend ou non des soins du visage humains. Les variations d'éclairage et l'expression du visage peuvent bloquer la détection

du visage. Afin de faciliter la conception d'un système de reconnaissance faciale robuste et de le rendre plus efficace, des étapes de prétraitement sont effectuées. De nombreuses approches sont utilisées pour détecter et définir l'image du visage humain, par exemple le détecteur Viola-Jones [32, 37], histogramme du gradient orienté (HOG) [26], et analyse en composantes principales (PCA) [29]. En outre, l'étape de détection de visage peut être utilisée pour la classification d'images et de vidéos [24], regression [9], suivi d'objets [30], détection de la région d'intérêt [29], etc.

- **Extraction de caractéristiques:** La fonction essentielle de cette étape est de décrire les images faciales capturées lors de l'étape de détection. Cette étape explique un visage comme un vecteur de groupe de traits appelé «traits» qui caractérise les traits discriminants de l'image du visage tels que la bouche, le nez et les yeux avec leur distribution géométrique [24]. Chaque visage est décrit par sa taille, sa structure et sa forme, ce qui lui permet d'être identifié. De nombreuses approches impliquent l'extraction de la forme / forme des yeux, de la bouche ou du nez pour identifier le visage en utilisant la taille et / ou la distance [24]. HOG [26], Eigenface [31], analyse indépendante des composants (ICA) [18], Filtre Gabor [20] les approches sont largement utilisées pour extraire les traits du visage.

- **Reconnaissance de visage:** Cette étape considère le vecteur de caractéristiques extrait de l'arrière-plan au cours de l'étape d'extraction de caractéristiques et le compare avec un visage similaire stocké dans un ensemble de données spécifique. Il existe deux applications générales essentielles pour la reconnaissance faciale, l'une est appelée reconnaissance ou identification et une autre est appelée vérification. Au cours de la phase d'identification, un test facial est comparé à un groupe de soins du visage visant à trouver la correspondance la plus similaire. Au cours de la phase de vérification, un facial de test est comparé à un facial connu de l'ensemble de données afin de prendre une décision d'acceptation ou de rejet.

Plusieurs systèmes proposés et mis en œuvre pour identifier un visage humain en images 2D ou 3D. Nous classons ces systèmes en trois méthodes en fonction de leur approche de détection et de reconnaissance (Fig. 1.3): (1) approches locales, (2) holistiques (sous-espace) et (3) hybrides. La première méthode est classée en fonction des caractéristiques spécifiques du visage, sans tenir compte de l'ensemble du visage. La deuxième méthode utilise le visage entier comme information d'entrée et se projette ensuite dans un sous-espace petit et discriminant ou dans un sous-plan de corrélation. La troisième méthode utilise des caractéristiques globales et locales afin d'améliorer la précision de la reconnaissance faciale.

La vérification de la parenté à partir d'images de visage, l'un des nouveaux sujets de la vision par ordinateur qui a été étudié et utilisé depuis plusieurs années, peut s'appliquer à

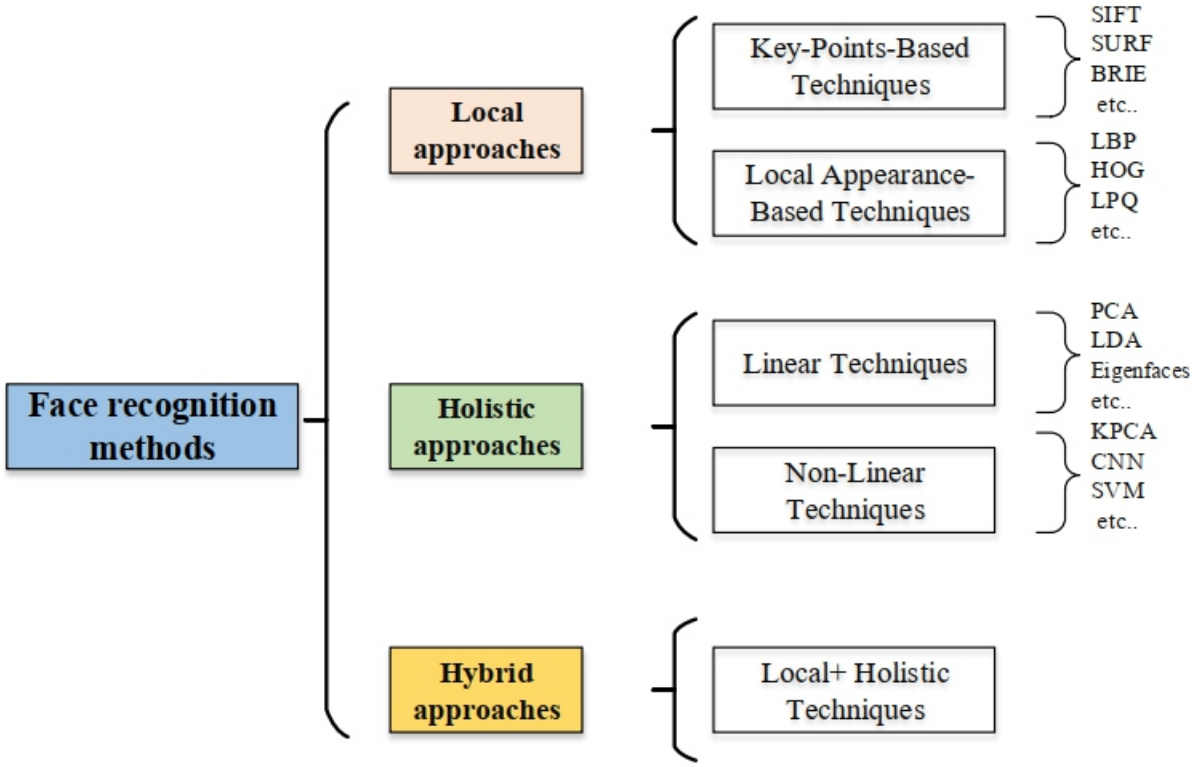


Figure 1.3: Méthodes de reconnaissance faciale. SIFT, transformation de caractéristique invariante d'échelle; SURF, transformation de caractéristique invariante d'échelle; BREF, caractéristiques élémentaires indépendantes robustes binaires; LBP, modèle binaire local; HOG, histogramme des dégradés orientés; LPQ, quantification de phase locale; PCA, analyse en composantes principales; LDA, analyse discriminante linéaire; KPCA, noyau PCA; CNN, réseau neuronal convolutif; SVM, supporte la machine vectorielle.

des applications potentielles, telles que la création d'arbres généalogiques, l'organisation d'albums de famille, l'annotation d'images, la recherche d'enfants disparus et la médecine légale. Vérifier si deux personnes appartiennent ou non à la même famille peut être automatiquement vérifié par des images faciales. Apprendre et extraire les similitudes de visage entre les membres de la famille est un défi. De nombreux résultats encourageants ont été obtenus au cours des dernières années, la vérification de la parenté à partir des images de visage reste ouverte. Bien qu'un test ADN soit le moyen le plus fiable pour la vérification de la parenté, il ne peut pas être utilisé dans de nombreuses situations. La vérification automatique de la parenté à partir d'images faciales peut être réalisée à titre d'exemple dans les scènes de vidéosurveillance. En plus de l'obstacle généralement confronté à la vérification du visage dans des environnements non contraints (c'est-à-dire des images faciales capturées dans des environnements non contrôlés sans aucune restriction en termes de pose, d'éclairage, d'arrière-plan, d'expression et d'occlusion partielle), la vérification de parenté insère une autre couche d'obstacles qui est loin d'être facile. La vérification de la parenté traite des images faciales qui appartiennent inévitablement à des personnes différentes avec une différence d'âge considérable et dans certaines conditions avec un sexe différent. En outre, les traits de visage des personnes de la même famille peuvent offrir une grande dissemblance alors que les visages de paires de personnes sans parenté peuvent

sembler similaires. Tous ces défis augmentent considérablement les difficultés du problème de la vérification automatique de la parenté.

À travers les différents chapitres, nous mettons en évidence l'intérêt d'utiliser des algorithmes basés sur l'analyse mono-dimensionnelle (vectorielle) et multidimensionnelle (basée sur les tenseurs) utilisant des fonctionnalités profondes et superficielles dans la vérification de parenté.

1.2 Défis de vérification de la parenté

La vérification de la parenté, l'un des sujets de base de la vision par ordinateur et de la reconnaissance des formes, a reçu une attention considérable ces dernières années. De nombreuses approches ont été proposées pour vérifier la parenté dans des environnements sans contrainte, alors que chacune de ces approches consiste à vérifier si deux personnes sont de la même famille ou non à travers des images faciales.

La vérification de la parenté au moyen d'images faciales est difficile en raison du degré élevé de variabilité des effets visibles tels que la différence génétique, la différence entre les sexes et l'écart d'âge. En bref, les deux facteurs suivants ont un impact majeur sur la résolution de problèmes:

- **Des défis uniques:** L'écart d'apparence dans le problème de vérification de la parenté est beaucoup plus grand que dans la configuration de vérification faciale traditionnelle (par exemple, en regardant deux images avec des sexes différents et des âges différents, et en vérifiant si ces deux sujets ont une relation entre le parent et l'enfant). De plus, les relations entre différents parents auront différents modèles de similitude. Ceux-ci peuvent poser des défis majeurs pour tous les systèmes de vérification de la parenté faciale.

- **Défis communs:** En raison des défis liés à la vérification des visages, l'apparence des visages en gros plan est sensible aux changements de divers facteurs, tels que les variations des expressions faciales, l'obstruction et la position. En outre, certains autres facteurs d'influence peuvent être présentés dans la scène réelle, comme l'éclairage, l'opacité ou la faible résolution, peuvent changer la représentation visuelle de la parenté des soins du visage de diverses manières.

Figure 1.4 illustre les défis mentionnés. Plusieurs algorithmes ont été proposés pour relever ces défis au cours de la dernière décennie. Plus récemment, [1, 6], ont étudié diverses approches représentatives de vérification de la parenté dans la situation du petit échantillon de données de apprentissage uniquement avec quelques types de relations, à savoir père-fils, père-fille, mère-fils et mère-fille. Par conséquent, il reste à tester si les approches modernes de vérification de la parenté fonctionnent bien sur de grands

échantillons de données d'entraînement avec diverses relations plus étroites, en particulier maintenant que l'ensemble de données FIW publié a été publié. [27]. Comme l'illustre la figure 1.4, les relations d'échantillons plus variées des membres de la famille posent des défis plus importants à la question de la vérification de la parenté et sont loin d'être résolues.

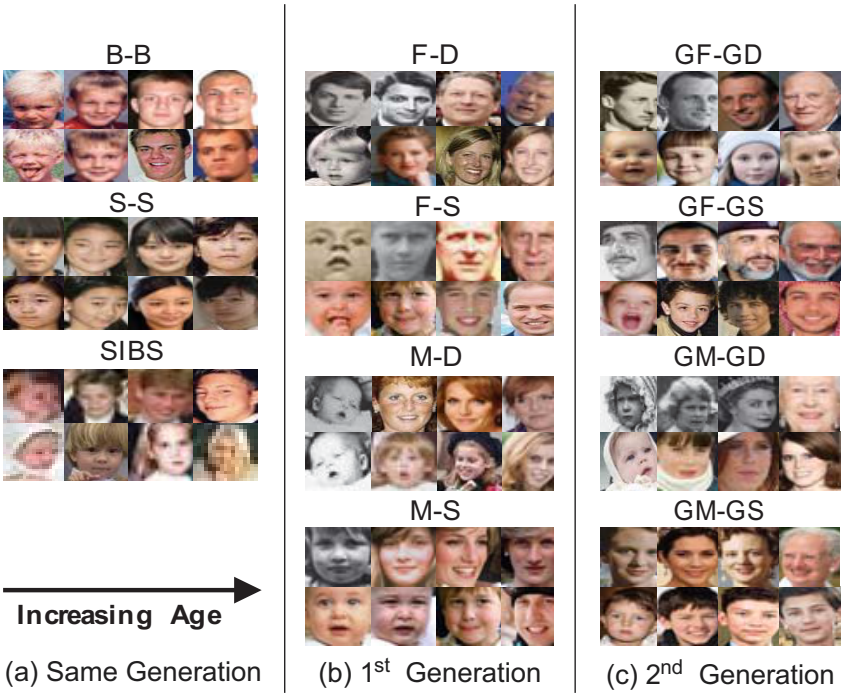


Figure 1.4: Échantillons de 11 types de paires de FIW. Chaque type est d'une paire unique sélectionnée au hasard dans un ensemble de familles diverses pour montrer la variation de l'appartenance ethnique, tandis que quatre visages de chaque individu représentent les variations d'âge [27].

1.3 Bases de données de référence

Pour évaluer les performances des approches de vérification de parenté proposées, nous avons considéré six bases de données de parenté: base de données Cornell KinFace, base de données UB KinFace, base de données TSKinFace, base de données KinFaceW-I, base de données KinFaceW-II et base de données FIW. Ces bases de données se composent de quatre types de relations parents-enfants (à l'exception de la base de données FIW qui contient onze relations de quatre relations parents-enfants, trois relations frères et sœurs et quatre relations grands-parents-petits-enfants). Les images de visage sont d'âges et d'ethnies variés, et capturées dans des environnements non contrôlés et sans restriction en termes de pose. Pour la vérification des visages (reconnaissance faciale ou auto-parenté), nous avons considéré deux bases de données complexes, à savoir la base de données Labeled Faces in the Wild (LFW) et la base de données YouTube Face (YTF).

1.3.1 Bases de données de vérification de parenté

Base de données Cornell KinFace [8] se compose de 143 paires d'images de parents et d'enfants recueillies sur le Web. Il y a 286 images de visage frontal recadrées de taille 100×100 pixels. La plupart des images ont été tirées de Google Images. Pour garantir que les caractéristiques du visage extraites sont de haute qualité, seules les images de face frontales avec une expression faciale neutre sont choisies. Nous notons que 7 familles sont retirées de la base de données originale qui se compose de 150 familles pour des questions de confidentialité.

Base de données UB KinFace [36] comprend 600 images de 400 personnes réparties en 200 paires de parents enfants-jeunes (ensemble 1) et 200 paires d'enfants-parents âgés (ensemble 2). Ces deux ensembles de paires sont utilisés pour améliorer, tester et évaluer les algorithmes de vérification de la parenté. La plupart des images de la base de données sont des combinaisons réelles de personnalités publiques (célébrités et politiciens) sur Internet. C'est la première base de données qui comprend tous les enfants, les jeunes parents et les vieux parents aux fins de vérification de la parenté.

Base de données TSKinFace [25] Il s'agit de deux types de relations de parenté tri-sujets qui sont: Père-Mère-Fille (FM-D) et Père-Mère-Fils (FM-S). Le FM-D contient 502 relations et FM-S a 513 relations (4060 images de visage). Ces images proviennent de personnalités publiques recueillies sur Internet. Les images du visage sont recadrées en utilisant la position des yeux dans 64×64 résolution de pixels. Pour une comparaison équitable, nous avons restructuré la base de données en séparant le groupe Père-Mère-Fille en deux groupes de relations de parenté Père-Fille et Mère-Fille, et le groupe Père-Mère-Fils en deux groupes Père-Fils et Mère-Fils relations de parenté.

Base de données visage de parenté dans la nature (KinFaceW) [19] se compose de deux sous-bases de données différentes: KinFaceW-I et KinFaceW-II. Les deux sous-bases de données sont rassemblées grâce à des recherches sur Internet, y compris des personnalités publiques avec leurs parents et / ou leurs enfants. Dans l'ensemble de données KinFaceW-I, il y a 156, 134, 116 et 127 paires correspondant aux relations F-S, F-D, M-S et M-D, respectivement. Pour l'ensemble de données KinFaceW-II, chaque type de relation parentale contient 250 paires. Au total, KinFaceW-I compte 1066 images de visage et 2000 images de visage pour KinFaceW-II.

Base de données FIW [27] nous avons considéré la plus grande base de données de parenté FIW utilisant: quatre sous-ensembles de visages Grand-père-Petite-fille (GF-GD), Grand-père-Petit-fils (GF-GS), Grand-mère-Petite-fille (GM-GD) et Grand-mère-Petit-fils (GM-GS). Dans le sous-ensemble GF-GD, il y a 7078 paires d'images pour les relations positives et négatives. Dans le sous-ensemble GF-GS, il y a 4830 paires d'images pour les relations positives et négatives. Dans le sous-ensemble GM-GD, il y a 6512 paires d'images pour les relations positives et négatives. Dans le sous-ensemble GM-GS, il y a 4614 paires

d'images pour les relations positives et négatives.

1.3.2 Bases de données de vérification faciale

Base de données des visages étiquetés dans la nature (LFW) [12] iest un grand ensemble de données collectées sur le Web, spécialement rassemblées pour étudier le problème de la reconnaissance faciale dans des environnements sans contraintes contenant des variations du monde réel en termes d'éclairage, de pose, d'expressions, de flou, d'occlusion, de résolution, etc. Cet ensemble de données difficile comprend 13233 images faciales appartenant à 5749 sujets différents.

Base de données YouTube Face (YTF) [35] se compose de 3425 vidéos de 1595 sujets différents avec diverses variations de pose, d'expression et d'illumination, et la longueur moyenne de chaque clip vidéo est de 181,3 images.

1.4 Objectifs et contributions

La thèse principale se concentre sur le développement, la mise en œuvre et l'évaluation de systèmes de vérification de parenté automatiques et efficaces basés sur des techniques d'apprentissage métrique de sous-espaces linéaires et multi-linéaires dans des environnements non contrôlés dans lesquels les variations de pose, d'éclairage, d'arrière-plan, d'expression et d'occlusion partielle sont très différent entre la apprentissage et les classes de test.

Nous pouvons organiser nos contributions en phase de classification en deux catégories essentielles: i) les méthodes vectorielles et ii) les méthodes tensorielles. Ces deux catégories ont besoin de l'étape d'extraction des caractéristiques, que nous avons classée par thème en deux catégories essentielles: i) les caractéristiques peu profondes (caractéristiques de forme/texture) et les caractéristiques profondes et superficielles. Figure 1.5 ont illustré nos principales contributions sur l'étape de classification en utilisant différentes catégories d'extraction de caractéristiques pour la vérification de la parenté.

Nous résumons les principales contributions de recherche de cette thèse comme suit:

- Etude de l'état de l'art (SOA) de différentes approches de vérification de parenté basées sur l'apprentissage profond et l'apprentissage métrique.
- Développer et concevoir des systèmes robustes de vérification de la parenté contre les variations d'expression, d'illumination et de pose, basés sur un apprentissage métrique vectoriel et basé sur les tenseurs, en utilisant les caractéristiques superficielles (c.-à-d. Texture/forme) et les caractéristiques profondes des images faciales d'intensité et des images faciales en couleur .

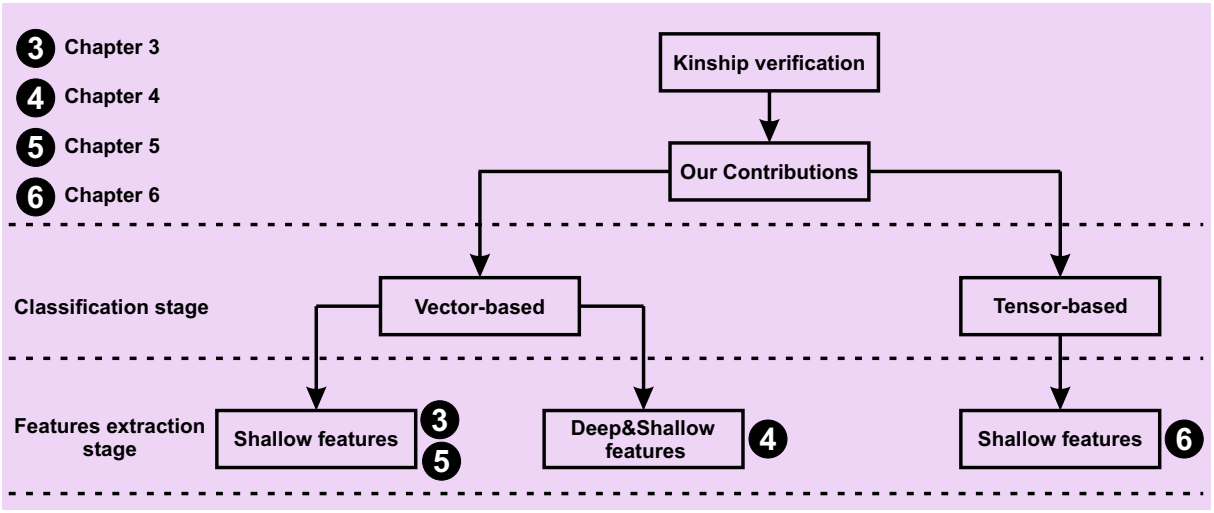


Figure 1.5: Carte de thèse: nos principales contributions sur différentes catégories d'étapes de classification (conception de stratégies vectorielles et basées sur des tenseurs) pour la vérification de la parenté en utilisant différentes catégories d'extraction de caractéristiques.

Nous pouvons subdiviser nos contributions en quatre volets comme suit:

I- Première contribution (illustré au Chapitre 3):

- Nous introduisons une méthode efficace pour la vérification de la parenté faciale basée sur l'extraction de caractéristiques à plusieurs échelles projetée à travers le sous-espace SIEDA (Side Information Exponential Discriminant Analysis) et combinant différentes caractéristiques à l'aide de la fusion des scores de la méthode de régression logistique (LR).

- Nous évaluons l'efficacité des données d'informations couleur-texture sur un sous-espace discriminant à l'aide d'une technique d'apprentissage en deux étapes, SIEDA et régression logistique, pour la vérification faciale automatique de la parenté à partir d'images faciales.

- Nous évaluons différents espaces colorimétriques et descripteurs sur quatre bases de données de parenté de référence. En particulier, chaque canal de couleur d'image de visage à partir d'un espace de couleur spécifié est projeté à travers le même sous-espace de canal de couleur appris implicite, puis toutes les informations de canal sont combinées pour obtenir une meilleure discrimination.

- Nous étudions la combinaison des différents descripteurs des différentes composantes de couleur.

II- Deuxième contribution (illustré au Chapitre 4):

- Nous introduisons un nouveau sous-espace discriminant de l'analyse discriminante linéaire basée sur les informations secondaires, intégrant la méthode d'analyse de

transformation de sous-espace de normalisation de covariance de classe (SILD+WCCN) pour la vérification de la parenté faciale. Par conséquent, le WCCN minimise l'impact intra-variabilité de classe en minimisant l'erreur de classification attendue au étape de l'apprentissage. [2].

- Nous suggérons deux systèmes de vérification de la parenté faciale automatisés robustes appropriés pour la vérification de la parenté bi-sujets et tri-sujets, à partir d'images faciales capturées dans des environnements sans contraintes. Les données faciales sont illustrées comme une fonction de vue multiple basée sur la fusion de différentes caractéristiques profondes et peu profondes afin d'obtenir un modèle facial plus discriminant.

- Nous évaluons l'efficacité des données d'informations profondes/superficielles sur un nouveau sous-espace discriminant en utilisant la technique d'apprentissage en deux étapes, SILD+WCCN et la régression logistique, pour la vérification automatique de la parenté à partir d'images faciales.

- Nous testons en profondeur notre technique SILD+WCCN/LR par rapport aux approches de pointe en utilisant deux bases de données de parenté faciales difficiles, à savoir KinFaceW-II et TSKinFace.

III- Troisième contribution (illustré au Chapitre 5):

- Présentation d'une nouvelle fonctionnalité native pour décrire les photos du visage. Notre descripteur est basé sur les caractéristiques statistiques locales de l'image du visage et le descripteur BSIF original.

- Nous proposons une nouvelle méthode SIWEDA pour vérifier le visage et la parenté basée sur la méthode classique SIEDA. De plus, pour atténuer la variance interne de la classe, nous avons proposé deux variantes SIEDA + WCCN et SIWEDA + WCCN en incorporant WCCN dans SIEDA et SIWEDA, respectivement.

- Nous évaluons globalement notre approche par rapport aux approches de pointe en utilisant cinq bases de données de visage et de parenté difficiles, à savoir Cornell KinFace, UB KinFace, TSKinFace, YTF et LFW.

IV- Quatrième contribution (illustré au Chapitre 6):

- Pour la première fois, nous traitons le problème de la vérification de la parenté faciale comme un problème de correspondance croisée car chaque parenté change généralement à partir de deux images faciales appartenant à deux personnes différentes.

- Nous suggérons un système de vérification automatique du visage robuste et adapté pour la vérification de la parenté, à partir de photos de visage prises dans des

environnements illimités. Les données de visage sont représentées sous la forme d'un tenseur de haut niveau qui repose sur une combinaison de différentes caractéristiques locales afin de fournir un modèle de visage plus robuste.

- Nous proposons une nouvelle méthode de réduction et de classification des dimensions, appelée analyse discriminante quadratique Tensor Cross-view (TXQDA), qui préserve la structure des données, élargit la marge entre les échantillons, aide à atténuer le problème de la petite taille des échantillons et réduit les coûts de calcul.

- Nous évaluons globalement notre méthode TXQDA par rapport à des méthodes de pointe en utilisant cinq bases de données de parenté faciales difficiles à savoir Cornell KinFace, UB KinFace, TSKinFace, KinFaceW-II et FIW.

Enfin, nous pouvons classer nos contributions (citées dans les chapitres 3,4,5 et 6) comme mentionné dans la figure 1.5 comme suit: Au Chapitre 3, nous avons utilisé la combinaison d'entités peu profondes projetées par la méthode SIEDA (basée sur les vecteurs). Au Chapitre 4, nous avons utilisé la combinaison des caractéristiques profondes et peu profondes projetées par la méthode proposée SILD+WCCN (basée sur les vecteurs). Pour le Chapitre 5, nous avons utilisé la combinaison d'entités peu profondes projetées par la méthode proposée SIWEDA+WCCN (basée sur les vecteurs). Pour le Chapitre 6, nous avons utilisé la combinaison de caractéristiques peu profondes sur une conception de tenseur projetée par la méthode proposée TXQDA (basée sur les tenseurs).

1.5 Explications d'apprentissage automatique

Dans le domaine de la vision par ordinateur, un système publié dans les travaux de la littérature doit prendre en compte plusieurs points et doit être explicable. Comme mentionné dans [22], une méthode d'explication pour une approche d'apprentissage automatique en boîte noire (un système / une méthode) devrait prendre en compte les propriétés suivantes:

• **Précision**. Ce trait fait référence au degré de succès d'une explication qui prédit de nouvelles données testées (données invisibles). une faible précision des explications ne peut convenir que si le système de la boîte noire à expliquer est également inexact.

• **Fidélité**. Les prédictions du modèle expliquées doivent correspondre et conclure les explications. Il existe une relation élevée entre la précision et la fidélité: lorsque l'explication a une plus grande fidélité et que le modèle de boîte noire est très précis, l'explication du modèle a également une plus grande précision.

• **Cohérence**. Les explications doivent s'appliquer de la même manière à tous les modèles entraînés à l'aide du même ensemble de données de train.

• **Stabilité**. Des instances similaires doivent présenter des explications similaires, tant que des instances particulières ont été fournies.

• **Représentativité**. Une explication très représentative est celle qui peut être appliquée à de nombreuses décisions dans de nombreux cas.

• **Certitude**. Si la méthode à l'étude fournit une mesure de confiance dans ses décisions, une explication de cette décision doit en tenir compte.

• **Nouveauté**. Cette propriété indique la capacité du paradigme d'explication à couvrir des instances éloignées de l'espace d'apprentissage.

• **Degré d'importance**. L'explication doit mettre en évidence les caractéristiques importantes.

• **Compréhensibilité**. Les explications doivent être compréhensibles pour les humains. Cela appartient au public cible et a des implications psychologiques et sociales, bien que de brèves explications contribuent généralement à la compréhension.

Miller a étudié l'explicabilité du point de vue des sciences sociales [21] et note quatre observations essentielles: (i) les gens donnent la priorité aux explications contrastives, c'est-à-dire pourquoi le modèle a pris une décision spécifique n'a pas autant d'importance pour nous que pourquoi une décision différente n'a pas été prise à la place; (ii) les gens ne choisissent que quelques raisons parmi les diverses raisons qui composent une explication, et les préjugés personnels témoignent de cette sélection; (iii) renvoyer aux probabilités ou aux liens statistiques n'est pas aussi efficace que renvoyer aux raisons; et (iv) les explications sont sociales et devraient donc faire partie d'une conversation plus large, ou d'une interaction entre l'explicateur et la personne expliquée.

In [11], les auteurs confirment l'importance des experts du domaine humain pour guider la croissance et l'évaluation des paradigmes d'explication, étant donné que les systèmes d'apprentissage automatique actuels fonctionnent sur un système statistique et / ou sans modèle, et exigent un contexte de systèmes humains / scientifiques pour transférer des explications convaincantes (en particulier pour les autres experts du domaine). Aucun modèle d'explication unique dans la littérature actuelle n'est en mesure de satisfaire toutes les propriétés mentionnées (pour plus de détails, reportez-vous à [3, 10, 22]pour les manuels d'enquêtes approfondies sur les systèmes d'intelligence artificielle explicables).

Dans nos systèmes, tous ces points ont été pris en considération. De plus, comme mentionné dans les Chapitres 3, 4, 5 et 6, nos contributions/publications ont atteint tous ces points.

1.6 Répercussion des limites de la biométrie sur les systèmes de parenté

Les systèmes de vérification de parenté souffrent et sont également affectés par les facteurs traditionnels des systèmes biométriques. Certains des principaux facteurs affectant la précision des systèmes biométriques [15] sont les suivants:

1. **Bruit dans les données détectées:** Le bruit dans l'échantillon de visage biométrique

obtenu peut résulter de caméras dégradées et mal entretenues ou de conditions ambiantes défavorables (environnement sans contrainte). Par exemple, une caméra de qualité peut également produire une image de visage bruyante comme le montre la figure 1.6. L'échantillon de visage biométrique bruyant ne pouvait pas être mis en correspondance, pour les utilisateurs authentiques, par leurs modèles compétents dans l'ensemble de données ou par le cloud malencontreux mis en correspondance avec les imposteurs, ce qui conduisait à une minimisation considérable des performances du système [33, 34].

Figure 1.6: systèmes typiques de vidéos de surveillance. (a) et (c) sont les images de surveillance d'une caméra de taille CIF (pixels) et d'une caméra de taille 720P (pixels) respectivement; (b) montre deux faces intéressées bruyantes extraites de (a) et (c) [17].

2. **Variations intra-classe:** Les variations intra-classe dans les échantillons de visage biométriques sont des exemples créés par l'interaction inappropriée de l'utilisateur avec l'intervalle de temps à la capture ou la caméra, par exemple, pose faciale incorrecte - voir Fig. 1.7, changements dans le conditions d'environnement (par exemple, changements d'éclairage) [38], utilisation de diverses caméras pendant l'inscription et la vérification, ou variation temporelle des caractéristiques biométriques telles que le vieillissement [7]. D'énormes variations intra-classe réduisent généralement le taux d'acceptation réel (GAR) d'un système biométrique.

3. **Similitudes inter-classes:** La similarité inter-classes est connue sous le nom d'interférence des échantillons biométriques, dans l'espace au niveau des entités, selon différentes classes ou peuples. La faiblesse de la singularité dans l'ensemble des traits biométriques conduisant à une maximisation du taux de fausse acceptation (FAR) du système. Par conséquent, il y a une limite maximale sur le nombre d'individus singuliers qui pourraient être assimilés par le système biométrique.

4. **Non-universalité:** L'universalité indique que chaque personne utilisant un système biométrique est capable de préparer la caractéristique biométrique respective. Le système biométrique ne pouvait pas être capable d'extraire des informations biométriques significatives d'un ensemble de données d'utilisateurs. Par exemple,

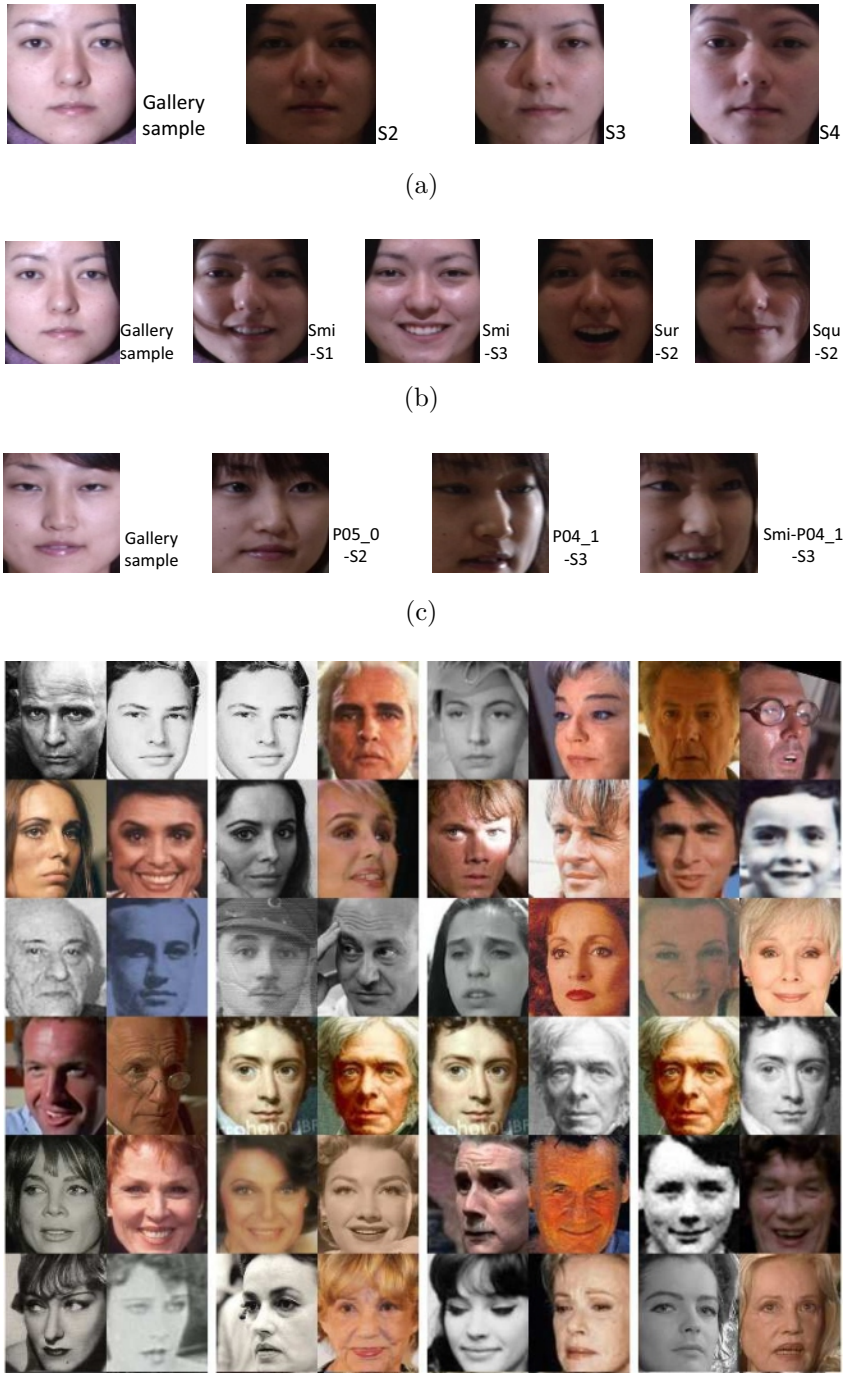


Figure 1.7: Exemples d'images de visage avec (a) des variations d'éclairage dans différentes sessions [38], (b) des variations d'expression dans différentes sessions [38], (c) posent des variations dans différentes sessions [38], et (d) Des paires d'échantillons positifs d'AgeDB [23] avec un écart de 30 ans, les apparences faciales subissent des changements dramatiques au cours de cette période [7].

l'Institut national des normes et de la technologie (NIST) a signalé qu'il est impossible d'extraire les traits de minutie corrects des empreintes digitales de deux personnes de la population (travailleurs manuels avec diverses ecchymoses et coupures au bout des doigts, individus avec incapacités liées etc.), en raison de la faible qualité des crêtes [5]. Cela contribue à maximiser le taux d'échec d'inscription (ETP). Par

conséquent, il n'y a pas de fonction biométrique qui est vraiment universelle.

5. **Problèmes d'interopérabilité:** Généralités, les systèmes biométriques sont contagieux et conçus par la présomption que l'échantillon biométrique à vérifier est acquis en utilisant la même caméra et, par conséquent, sont limités dans leur capacité à correspondre ou à vérifier les échantillons biométriques résultant de diverses caméras.

6. **Attaques par usurpation:** Une attaque par usurpation biométrique est la tentative intentionnelle de falsifier ses caractéristiques biométriques afin d'éviter la vérification, ou l'induction d'artefacts biométriques physiques afin de se réincarner sur l'identité d'un autre individu.

   Dans le domaine de la vérification de parenté, toutes les méthodes proposées doivent être confrontées à tous ces défis (défis des systèmes biométriques) avant de traiter le facteur de parenté. Dans les Chapitres 3, 4, 5 et 6, Nos résultats réalisés par les systèmes proposés montrent que tous ces défis sont traités avec succès sur des bases de données de vérification de parenté capturées dans des environnements sans contraintes.

1.7 Évaluation des performances

La biométrie est le terme scientifique pour les mesures et les calculs corporels. Cela indique que les métriques appartenaient à des traits humains. La vérification biométrique (ou authentification réaliste) est utilisée dans la vision par ordinateur comme une forme de correspondance d'identité et de contrôle d'accès. Nous nous référons à cela également utilisé pour identifier les personnes dans les groupes qui sont sous surveillance.

Les identificateurs biométriques sont les traits discriminatoires et mesurables utilisés pour décrire et étiqueter les personnes. Les reconnaisseurs biométriques sont très majoritairement classés comme physiologiques par rapport aux traits comportementaux. Les traits physiologiques appartiennent à la forme du corps. Les exemples impliquent, mais sans s'y limiter, les veines de la paume, l'empreinte digitale, la reconnaissance faciale, l'empreinte de la paume, l'ADN, la géométrie de la main, la rétine et l'odeur / l'odeur, la reconnaissance de l'iris. Les traits de comportement appartiennent à la modalité de comportement d'un individu, impliquant, mais sans s'y limiter, le rythme de frappe, la démarche et la voix. Certains chercheurs ont utilisé le terme de comportement pour caractériser cette dernière classe de biométrie.

Des moyens plus conventionnels d'accès au contrôle comprennent des systèmes de reconnaissance basés sur des jetons, tels qu'un permis de conduire ou un passeport, et des systèmes d'identification basés sur des connaissances, tels qu'un mot de passe ou un numéro d'identification personnel. Étant donné que les identificateurs biométriques sont

uniques à chaque personne, mais qu'ils sont très fiables dans la vérification d'identité que les méthodes basées sur les jetons et les connaissances; en outre, l'ensemble d'identifiants biométriques soulève les problèmes de confidentialité concernant l'utilisation intégrale de ces caractéristiques d'information.

1.7.1 Fonctionnalité biométrique

De nombreux aspects divers de la physiologie, du comportement ou de la chimie d'une personne peuvent être utilisés pour la vérification biométrique. Le choix d'une biométrique spécifique à utiliser dans une application particulière comprend une pondération de divers facteurs. Jain et coll. [13] décrivent sept facteurs importants (points d'indication) à utiliser lors de l'estimation de l'aptitude de toute caractéristique à être utilisée dans la vérification biométrique.

- L'universalité signifie que chaque personne utilisant un système doit prendre possession de la fonctionnalité.
- Unicité signifie que le trait doit être suffisamment varié pour les personnes de la population pertinente de sorte qu'elles puissent être différenciées les unes des autres.
- La permanence appartient à la manière dont une caractéristique change au fil du temps. Plus particulièrement, une fonctionnalité avec une «bonne» permanence sera rationnellement inchangée au fil du temps avec conservation du modèle correspondant particulier.
- La mesurabilité (collectabilité) appartient à l'installation de mesure et / ou de stockage de l'élément. En outre, les informations acquises doivent être sous une forme permettant simplement le traitement et l'extraction ultérieurs des ensembles de caractères pertinents.
- Les performances appartiennent à la vitesse, la précision, la robustesse et l'efficacité de la technologie utilisée (voir la sous-section performance 1.7.2 pour plus de détails).
- L'acceptabilité dépend de la mesure dans laquelle les personnes de la population concernée consentent à la technologie dans laquelle elles sont prêtes à faire capturer et traiter leur caractéristique biométrique.
- Le contournement appartient au simple avec lequel une caractéristique doit être imitée en utilisant un substitut ou un artefact.

Une utilisation biométrique appropriée dépend extrêmement de l'application. En outre, certaines données biométriques devraient être meilleures que d'autres en fonction des

niveaux de sécurité demandés en termes de commodité de sécurité [4]. Aucun monoculaire biométrique ne doit répondre à toutes les exigences de toutes les applications possibles [13].

Le schéma bloc du système bimétrique comprend deux modes essentiels d'un modèle biométrique [14]. Premièrement, dans le modèle de vérification (ou d'authentification), le système exécute une vérification individuelle d'une mesure biométrique avec un modèle spécial stocké dans un ensemble de données biométriques dans lequel la vérification plus simple de la cible individuelle est la même personne qu'elle devrait être. En général, trois procédures sont incluses dans la vérification d'identité de la personne [28]. Dans un premier temps, des systèmes de référence pour tous les utilisateurs où ils sont produits et stockés dans le jeu de données-système. Dans la deuxième étape, certains des échantillons où ils sont appariés avec des systèmes de référence pour produire l'imposteur et les scores réels et calculer le seuil. La troisième étape est la phase de test. Cette étape peut utiliser une carte à puce (SC), un numéro d'identification ou un nom d'utilisateur (par exemple un PIN) pour signaler quel modèle doit être utilisé pour la comparaison de vérification. «L'identification positive» est une utilisation conjointe du mode d'authentification, «où le but est d'empêcher plusieurs utilisateurs d'utiliser l'identité unique» [28].

Deuxièmement, le mode d'identification / reconnaissance, le système procède à une vérification / comparaison un-à-plusieurs par rapport à un ensemble de données biométriques dans le but de trouver l'identité d'une personne inconnue. Le système pourrait réussir à gérer l'identification de la personne si la vérification de l'échantillon biométrique testé à un modèle dans l'ensemble de données tombe à l'intérieur d'un seuil préalablement fixé. Le mode de reconnaissance / identification peut être utilisé beaucoup pour la «vraie reconnaissance» (afin que l'utilisateur n'ait pas à afficher de données d'information sur le modèle à utiliser) ou pour la «fausse reconnaissance» de l'individu »où le système détermine si le l'individu est celui qu'elle nie (explicitement ou implicitement) être " [28]. Ce dernier ne peut être fait que sur la biométrie puisque les autres approches d'identification personnelle telles que les codes PIN, les mots de passe ou les clés sont inefficaces.

1.7.2 Performance

Dans ce qui suit, les utilisés comme mesures de performance pour les systèmes biométriques:

- **False match rate** (FMR, également appelé FAR = False Accept Rate): représente la probabilité que le framework ait mal classé le modèle d'entrée de test en un échantillon non correspondant dans l'ensemble de données. Il représente le pourcentage d'entrées nulles et non avenues qui sont mal acceptées. En situation de mesure de similarité, si l'individu est en réalité un imposteur, mais que le score apparié est supérieur au seuil, alors nous l'avons traité comme authentique. Ceci maximise le FMR, dans lequel repose donc également sur le score seuil [28].
- **Faux taux de non-correspondance** (FNMR, également nommé FRR = Faux taux

de rejet): la probabilité que le système indique à tort qu'il existe une correspondance entre le modèle d'entrée de l'échantillon et un modèle correspondant dans l'ensemble de données. Il calcule le pourcentage d'entrées utiles qui sont rejetées à tort.

- **Caractéristique de fonctionnement du récepteur** ou caractéristique de fonctionnement relative (ROC): Le ROC est une représentation graphique graphique du compromis entre le FMR (FAR) et le FNMR (FRR). En général, l'approche d'appariement génère une décision basée sur un seuil qui définit à quel point l'échantillon d'entrée doit nécessairement être proche d'un modèle pour qu'il ressemble à une correspondance. Lorsque le seuil est diminué, il doit y avoir le plus petit nombre de fausses non-concordances mais d'autres fausses acceptations. De plus, un seuil plus grand devrait diminuer le FMR mais élargir le FNMR. Une différence commune est le compromis d'erreur de détection (DET), par lequel il est acquis en utilisant des échelles d'écart normales sur les deux axes. Ce graphe plus linéaire allège les divergences pour de meilleures performances (erreurs rares).

- **Taux d'erreur égal** ou taux d'erreur de croisement (EER ou CER): le taux auquel les erreurs de rejet et d'acceptation sont égales. Le taux de l'EER peut être simplement extrait à l'aide de la courbe ROC. L'EER est un moyen rapide de comparer la précision des machines avec différentes courbes ROC. Généralement, la machine avec le plus petit EER est la haute précision.

- **Taux d'échec d'inscription** (FTE ou FER): le taux auquel tente de générer un modèle à partir d'un échantillon d'entrée échoue. Ceci est généralement inspiré par des entrées d'échantillons de faible qualité.

- **Taux d'échec de capture** (FTC): dans les systèmes automatiques, la probabilité que le système ne parvienne pas à déterminer une entrée d'échantillon biométrique lorsqu'il est donné correctement.

- **Capacité du modèle**: le nombre extrême de collections de données qui peuvent être stockées dans le framework.

1.8 Articulation de la thèse

Le manuscrit de thèse est structuré autour de sept chapitres:

Dans le Chapitre 1, nous avons donné une introduction générale des contextes, motivations, objectifs et contributions de cette thèse.

Dans le Chapitre 2, nous mentionnons un aperçu général des méthodes de pointe de la vérification de la parenté ainsi que de leurs différents types: fonctionnalités de vérification de la parenté basée sur l'apprentissage, la vérification de la parenté basée sur l'apprentissage

métrique et la vérification de la parenté convolutive basée sur l'apprentissage profond . D'autre part, nous avons présenté le problème de parenté à partir d'images faciales et ses caractéristiques de mesure ainsi que le système général de vérification de la parenté.

Dans le Chapitre 3 (notre première contribution), nous présentons une approche de vérification de parenté faciale (FKV) basée sur un apprentissage automatique et plus efficace en deux étapes de la couleur / informations de texture. La plupart des méthodes proposées pour la vérification automatique de la parenté à partir des images de visage ne prennent en compte que les informations de luminance (c'est-à-dire l'échelle de gris) et excluent les informations de chrominance (c'est-à-dire la couleur) qui peuvent être utiles, en tant qu'indice supplémentaire, pour prédire les relations. Nous explorons l'utilisation conjointe des informations couleur-texture de la chrominance et des canaux de luminance en extrayant des caractéristiques complémentaires de bas niveau à partir de différents espaces colorimétriques. Plus spécifiquement, les caractéristiques sont extraites de chaque canal de couleur de l'image du visage et fusionnées pour obtenir une meilleure discrimination. Nous étudions différents descripteurs sur les bases de données de parenté de visage existantes, illustrant l'utilité des informations de couleur, par rapport aux homologues en échelle de gris, dans sept espaces colorimétriques différents. En particulier, nous générons à partir de chaque espace colorimétrique trois matrices de projection de sous-espaces, puis évaluons la méthodologie de fusion pour fusionner trois distances appartenant à chaque image de visage de paire de test. Des expériences sur trois bases de données de référence, à savoir Cornell KinFace, KinFaceW (I & II) et la base de données TSKinFace, montrent des résultats supérieurs par rapport à l'état de l'art.

Dans le Chapitre 4 (notre deuxième contribution), nous présentons la combinaison de caractéristiques profondes et superficielles (caractéristiques multi-vues) en utilisant l'approche d'apprentissage métrique proposée (SILD + WCCN / LR) pour la vérification de la parenté. Notre approche basée sur un apprentissage automatique et plus efficace en deux étapes de l'information profonde / superficielle. Tout d'abord, cinq couches pour les traits profonds et cinq traits superficiels (c.-à-d. Texture et forme), représentant plus précisément les traits du visage impliqués dans les relations de parenté (père-fils, père-fille, mère-fils et mère-fille) sont utilisées pour former le Proposition d'analyse discriminante linéaire basée sur les informations secondaires intégrant la méthode de normalisation de la covariance de classe (SILD + WCCN). Ensuite, chacune des caractéristiques projetées à travers le sous-espace discriminant de la méthode d'apprentissage métrique SILD + WCCN proposée. Enfin, une méthode de régression logistique (LR) est utilisée pour fusionner les six scores des caractéristiques projetées. Pour montrer l'efficacité de notre méthode SILD + WCNN, nous faisons quelques expériences sur la base de données LFW. En termes d'évaluation, la vérification automatique de la parenté faciale (FKV) proposée est comparée à celles existantes pour montrer son efficacité, en utilisant deux bases de données de parenté difficiles. Les résultats expérimentaux ont montré la supériorité de

notre FKV par rapport à ceux existants pour l'appariement bi-sujet sur KinFaceW-II et TSKinFace. De plus, les résultats expérimentaux ont montré la supériorité de notre FKV sur la base de données TSKinFace disponible pour Père-Mère-Fils et Père-Mère-Fille.

Dans le Chapitre 5 (notre troisième contribution), nous développons un nouveau critère, appelé Analyse Discriminante Exponentielle Pondérée basée sur les informations secondaires (SIWEDA), basé sur la méthode classique SIEDA. Nous reformulons et généralisons la fonction de critère de Fisher classique afin de la maximiser, avec la propriété de rapprocher le plus possible les échantillons intra-classes (échantillons intra-classes), et de repousser et repousser le plus loin possible les échantillons inter-classes (échantillons inter-classes). Ainsi, SIWEDA sélectionne les valeurs propres de haute signification et élimine celles avec des informations moins discriminantes. Pour réduire la dimensionnalité du vecteur de caractéristiques et alléger l'intra-variabilité de classe, nous utilisons SIWEDA et la normalisation de covariance intra-classe (WCCN) en utilisant les caractéristiques d'image binarisées statistiques proposées (StatBIF). De plus, nous utilisons la stratégie de fusion des scores pour extraire la complémentarité des différentes échelles de pondération de notre descripteur StatBIF. Nous avons mené des expériences pour évaluer les performances de la méthode proposée dans un environnement sans contrainte, en utilisant cinq ensembles de données à savoir LFW, YTF, Cornell KinFace, UB KinFace et TSKinFace, dans le contexte de la correspondance des visages et de la vérification de la parenté dans des conditions sauvages. Les expériences ont montré que l'approche proposée surpasse l'état actuel de la technique. Très intéressant, notre approche a montré des performances supérieures par rapport aux méthodes basées sur l'apprentissage métrique profond

Dans le Chapitre 6 (notre quatrième contribution), nous présentons une nouvelle méthode d'analyse discriminante quadratique Tensor Cross-view (TXQDA) basée sur la méthode XQDA pour la vérification de la parenté dans la nature. De nombreux chercheurs ont utilisé des méthodes d'apprentissage métrique et ont obtenu des performances raisonnablement bonnes dans la vérification de la parenté, aucune de ces méthodes ne considère la vérification de la parenté comme un problème de correspondance croisée. Pour résoudre ce problème, nous proposons une méthode de vue croisée des tenseurs pour former des données multilinéaires à l'aide d'histogrammes locaux de descripteurs de caractéristiques locales. Par conséquent, nous apprenons une transformation tensorielle hiérarchique pour projeter chaque paire d'images de visage dans le même espace de caractéristiques implicite, dans lequel la distance de chaque paire positive est minimisée et celle de chaque paire négative est maximisée. De plus, TXQDA a été proposé de séparer la structure multifactorielle des images de visage (c'est-à-dire la parenté, l'âge, le sexe, l'expression, l'illumination et la pose) des différentes dimensions du tenseur. Ainsi, notre TXQDA obtient de meilleurs résultats de classification en découvrant un sous-espace tenseur de faible dimension qui agrandit la marge de différentes classes de relations de parenté. L'évaluation expérimentale de cinq bases de données complexes, à savoir Cornell KinFace, UB KinFace, TSKinFace,

KinFaceW-II et FIW, montre que le TXQDA proposé surpasse considérablement l'état actuel de la technique. De plus, notre méthode TXQDA fonctionne bien sur les classes de données d'entraînement les plus petites ou les plus limitées et sur les classes de données d'entraînement les plus grandes ou à grande échelle.

Dans le Chapitre 7, nous concluons cette thèse en résumant les principaux points de nos contributions et nous mentionnons quelques perspectives intéressantes à explorer à la suite de nos travaux.

Bibliographie

[1] Almuashi, M., Hashim, S.Z.M., Mohamad, D., Alkawaz, M.H., Ali, A.: Automated kinship verification and identification through human facial images: a survey. Multimedia Tools and Applications **76**(1), 265–307 (2017)

[2] Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: 2013 IEEE International Conference on Computer Vision, pp. 1960–1967 (2013). DOI 10.1109/ICCV.2013.246

[3] Biran, O., Cotton, C.: Explanation and justification in machine learning: A survey. In: IJCAI-17 workshop on explainable AI (XAI), vol. 8, pp. 8–13 (2017)

[4] Bleicher, P.: Biometrics comes of age: Despite accuracy and security concerns, biometrics are gaining in popularity. Applied Clinical Trials **14**(12), 18–20 (2005)

[5] Carmen, S.: Nist report to the united states congress. summary of nist tandards for biometric accuracy, tamper resistance and interoperabiity [r/ol] (2001)

[6] Dandekar, A.R., Nimbarte, M.S.: A survey: Verification of family relationship from parents and child facial images. In: 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, pp. 1–6 (2014)

[7] Deng, J., Zhou, Y., Zafeiriou, S.: Marginal loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 60–68 (2017)

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

[9] Guehairia, O., Ouamane, A., Dornaika, F., Taleb-Ahmed, A.: Feature fusion via deep random forest for facial age estimation. Neural Networks **130**, 238–252 (2020)

[10] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM computing surveys (CSUR) **51**(5), 1–42 (2018)

[11] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **9**(4), e1312 (2019)

[12] Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: 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`

[13] Jain, A.K., Bolle, R., Pankanti, S.: Personal identification in networked society. Biometrics (1999)

[14] Jain, A.K., Flynn, P., Ross, A.A.: Handbook of biometrics. Springer Science & Business Media (2007)

[15] Jain, A.K., Ross, A.: Multibiometric systems. Communications of the ACM **47**(1), 34–40 (2004)

[16] Jain, A.K., Ross, A.A., Nandakumar, K.: Introduction to biometrics. Springer Science & Business Media (2011)

[17] Jiang, J., Hu, R., Wang, Z., Han, Z.: Noise robust face hallucination via locality-constrained representation. IEEE Transactions on Multimedia **16**(5), 1268–1281 (2014)

[18] Lee, T.W.: Independent component analysis. In: Independent component analysis, pp. 27–66. Springer (1998)

[19] Lu, J., Zhou, X., Tan, Y.P., Shang, Y., Zhou, J.: Neighborhood repulsed metric learning for kinship verification. IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 331–345 (2014). DOI 10.1109/TPAMI.2013.134. URL `http://dx.doi.org/10.1109/TPAMI.2013.134`

[20] Mehrotra, R., Namuduri, K.R., Ranganathan, N.: Gabor filter-based edge detection. Pattern recognition **25**(12), 1479–1494 (1992)

[21] Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence **267**, 1 – 38 (2019). DOI https://doi.org/10.1016/j.artint.2018.07.007. URL `http://www.sciencedirect.com/science/article/pii/S0004370218305988`

[22] Molnar, C.: Interpretable machine learning: a guide for making black box models explainable. leanpub (2018)

[23] Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S.: Agedb: the first manually collected, in-the-wild age database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 51–59 (2017)

[24] Ouamane, A., Boutellaa, E., Bengherabi, M., Taleb-Ahmed, A., Hadid, A.: A novel statistical and multiscale local binary feature for 2d and 3d face verification. Computers & Electrical Engineering **62**, 68–80 (2017)

[25] Qin, X., Tan, X., Chen, S.: Tri-subject kinship verification: Understanding the core of a family. IEEE Transactions on Multimedia **17**(10), 1855–1867 (2015). DOI 10.1109/TMM.2015.2461462

[26] Rettkowski, J., Boutros, A., Göhringer, D.: Hw/sw co-design of the hog algorithm on a xilinx zynq soc. Journal of Parallel and Distributed Computing **109**, 50–62 (2017)

[27] Robinson, J.P., Shao, M., Wu, Y., Liu, H., Gillis, T., Fu, Y.: Visual kinship recognition of families in the wild. In: IEEE Transactions on pattern analysis and machine intelligence (2018)

[28] Sahoo, S.K., Choubisa, T., Prasanna, S.M.: Multimodal biometric person authentication: A review. IETE Technical Review **29**(1), 54–75 (2012)

[29] Seo, H.J., Milanfar, P.: Face verification using the lark representation. IEEE Transactions on Information Forensics and Security **6**(4), 1275–1286 (2011)

[30] Touil, D.E., Terki, N., Medouakh, S.: Learning spatially correlation filters based on convolutional features via pso algorithm and two combined color spaces for visual tracking. Applied Intelligence **48**(9), 2837–2846 (2018)

[31] Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of cognitive neuroscience **3**(1), 71–86 (1991)

[32] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1, pp. I–I. IEEE (2001)

[33] Wilson, C., Hicklin, A., Bone, M., Korves, H., Grother, P., Ulery, B., Micheals, R., Zoepfl, M., Otto, S., Watson, C.: Fingerprint vendor technology evaluation 2003: summary of results and analysis report. NIST Technical Report NISTIR **7123** (2004)

[34] Wilson, C.L., Garris, M.D., Watson, C.I.: Matching performance for the us-visit ident system using flat fingerprints. In: National Institute of Standards and Technology Internal Report 7110. Citeseer (2004)

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

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

[37] Yang, H., Wang, X.A.: Cascade classifier for face detection. Journal of Algorithms & Computational Technology **10**(3), 187–197 (2016)

[38] Yang, M., Van Gool, L., Zhang, L.: Sparse variation dictionary learning for face recognition with a single training sample per person. In: Proceedings of the IEEE international conference on computer vision, pp. 689–696 (2013)