

Université Mohamed Khider Biskra
Faculté des Sciences et de la Technologie
Département : Génie Electrique
Ref :.....



جامعة محمد خيضر
بسكرة

جامعة محمد خيضر بسكرة
كلية العلوم و التكنولوجيا
قسم: الهندسة الكهربائية
المرجع:.....

Thèse présentée en vue de l'obtention
du diplôme de
Doctorat LMD en Génie Electrique

Filière : Electronique

Option : Signaux et Communication

**Reconnaissance de Visage basée sur l'Analyse
Multidimensionnelle**

Présentée par :

BESSAOUDI Mohcene

Soutenue publiquement le : 27/06/2019

Devant le jury composé de :

DEBILOU Abderrazak	Professeur	Président	Université de Biskra
BELAHCENE Mébarka	Professeur	Rapporteur	Université de Biskra
OUAMANE Abdelmalik	M. C. A	Co-Rapporteur	Université de Biskra
KAZAR Okba	Professeur	Examineur	Université de Biskra
SAMAI Djamel	M. C. A	Examineur	Université de Ouargla
MERAOUZIA Abdallah	M. C. A	Examineur	Université de Tebessa

Dédicaces

*A tout respect et amour je dédie ce travail à mes chers
parents, ma famille et tous mes amis*

Mohcene

Remerciements

Tout d'abord, je tiens à exprimer mes remerciements à ma directrice de thèse Mme **BELAHCENE** Mébarka, Professeur à l'Université Mohamed **KHIDER** Biskra, de m'avoir soutenu et fait confiance durant ces années avec une grande patience. Avec son expérience dans la recherche et l'enseignement, avec ses conseils, j'ai pu découvrir le monde de la recherche scientifique dans le domaine du traitement d'image et des techniques de la biométrie faciale. Il est effectivement un grand plaisir pour moi de travailler sous la supervision de Mme **BELAHCENE** Mébarka.

Je tiens à remercier chaleureusement Mr **OUAMANE** Abdelmalik, enseignant chercheur au laboratoire d'identification, de commande, contrôle et communication Biskra, pour sa contribution, son aide et ses encouragements qu'il m'a prodigués durant la période de la thèse.

Merci, bien évidemment à Mr **BOURENNANE** Salah Professeur à l'Ecole Centrale de Marseille en France et Responsable du GSM, Groupe Signaux Multidimensionnels pour son soutien, ses conseils fructueux ainsi que pour l'accueil dans son laboratoire.

Je tiens à remercier profondément les membres du jury :

1. Mr **DEBILOU** Abderrazak, Professeur à l'Université Mohamed **KHIDER** Biskra, trouve ici l'expression de mes remerciements les plus sincères pour avoir accepté de présider cette thèse.
2. Mr **KAZAR** Okba, Professeur à l'Université Mohamed **KHIDER** Biskra, pour l'intérêt porté à notre travail et sa participation au jury en tant qu'examineur.
3. Mr **SAMAI** Djamel, Maitre de Conférences A à l'Université Kasdi **MERBAH** Ouargla, d'avoir accepté d'être examinateur.
4. Mr **MERAOUIMIA** Abdallah, Maitre de Conférences A à l'Université Larbi **TEBESSI** Tebessa, d'avoir accepté d'être examinateur.

Enfin, je remercie mes amis et ma famille pour leur patience, leurs encouragements et leur compréhension concernant mon absence, surtout en fin de thèse.

Publications & Communications associées à la thèse

Les contributions présentées dans ce manuscrit de thèses ont été publiés dans les articles suivants:

I. Publications internationales

1. **M. Bessaoudi**, A. Ouamane, M. Belahcene, A. Chouchane, E. Boutellaa, and S. Bourennane, "Multilinear Side-Information based Discriminant Analysis for face and kinship verification in the wild," *Neurocomputing*, vol. 329, pp. 267-278, 2019. IF=3.241
2. **M. Bessaoudi**, M. Belahcene, A. Ouamane, A. Chouchane, and S. Bourennane, "Multilinear Enhanced Fisher Discriminant Analysis for robust multimodal 2D and 3D face verification," *Applied Intelligence*, vol. 49, pp. 1-16, 2018. IF=1.983

II. Communications internationales

1. **M. Bessaoudi**, M. Belahcene, A. Ouamane, A. Chouchane, and S. Bourennane, "A Novel Hybrid Approach for 3D Face Recognition Based on Higher Order Tensor," in *International Conference on Computer Science and its Applications*, 2018, pp. 215-224.
2. **M. Bessaoudi**, M. Belahcene, A. Ouamane, and S. Bourennane, "A novel approach based on high order tensor and multi-scale locals features for 3D face recognition," in *Advanced Technologies for Signal and Image Processing (ATSIP)*, 2018 4th International Conference on, 2018, pp. 1-5.
3. **M. Bessaoudi**, M. Belahcene, A. Ouamane, S. Bourennane, An Efficient Multimodal 2D and 3D Face Verification Based on Tensor Representation, 3rd International Conference on Pattern Analysis and Intelligent Systems, October 2018.

الملخص

في العقود الأخيرة، تلقى القياسات الحيوية الكثير من الاهتمام بسبب الحاجة المتزايدة باستمرار لمصادقة الهوية. ومن بين الطرائق البيومترية المختلفة، يقدم الوجه مزايا تنافسية لأن الحصول على بيانات الوجه أمر طبيعي، غير تدخلي ومقبول بشكل جيد من قبل البشر. استخدام تقنيات التعلم الفرعية متعددة الطبقات في التعرف على الوجه اثار اهتماما شديدا من قبل المجتمع العلمي للحد من الأبعاد وتصنيف البيانات متعددة الأبعاد. تهدف هذه الاطروحة الى التعرف التلقائي على الوجوه على أساس التحليل متعدد الأبعاد. في الجزء الأول، نقترح طريقة إسقاط جديدة متعددة الأبعاد، تسمى MEFDA (التحليل التعددي المحسن للتمييز فيشر) للتحقق من الوجه متعدد الوسائط (3D+2D). تعمل هذه الطريقة على تحسين قدرة التعميم عن طريق تحليل عملية التمييز إلى تقطير متزامن لمصفوفات نشيت الطبقة الداخلية والخارجية الناتجة عن نشر بيانات الموترات. بالإضافة إلى ذلك، تستخدم MEFDA معيار الموتر المميز بناءً على إجراء تكراري للحصول على مصفوفات الإسقاط المثلى في كل وضع الموتر. علاوة على ذلك، في هذه الطريقة، يتم احترام البنية الأصلية والهندسة الطبيعية للبيانات. تم تقييم الخوارزمية المقترحة على ثلاثة قواعد بيانات عامة للوجه ثلاثي الأبعاد، CASIA 3D، FRGC V2.0، وBosphorus، في السياق الصعب لتعبيرات الإضاءة وتعابير الوجه ودوران الرأس. تظهر النتائج التجريبية أن الطريقة المقترحة تتفوق بشكل كبير على الأساليب الأخرى. في الجزء الثاني، نقترح طريقة جديدة متعددة الأبعاد للتحقق من الوجه والقرابة في البيئات غير الخاضعة للتحكم. نسمي الخوارزمية الجديدة MSIDA (التحليل التعددي القائم على المعلومات الجانبية) ونستخدمها لحل مشكلة مطابقة زوج الوجه، والتي تتميز ببيانات ضعيفة. يمكن لهذه الطريقة الجديدة أن تمثل بشكل طبيعي صورًا متعددة الأبعاد واستخراج معلومات مفيدة مباشرة من بيانات الموتر بدلاً من المصفوفات أو المتجهات. في طريقة MSIDA، يتم الحصول على العديد من المساحات الفرعية المترابطة فيما بينها على أوضاع موتر مختلفة، بحيث يتم تعلم المساحات الجزئية بشكل متكرر عن طريق نشر الموتر في الأوضاع المختلفة. باستخدام معلومات ضعيفة، تقوم MSIDA بإنتاج موتر الوجه في فضاء جزئي جديد يتم فيه تحسين التمييز. يبين التقييم التجريبي لأربع قواعد بيانات الوجه (LFW، Cornell KinFace، UB KinFace و TSKinFace) أن المنهج المقترح يفوق بشكل كبير على الأساليب الحالية.

الكلمات المفتاحية: التعرف على الوجوه، تحليل متعدد الأبعاد، MEFDA، MSIDA، التعلم متعدد المراحل، التمثيل بواسطة الموترات.

Abstract

In the recent decades, biometrics has received a lot of attention because of the great need for identity authentication and identification. Among various biometric modalities, the face offers competitive advantages because the acquisition of facial data is natural, non-intrusive and well accepted by humans. The use of multilinear subspace learning techniques in face recognition has generated a great deal of interest from the scientific community. This thesis is devoted to automatic face recognition based on multidimensional analysis. In the first part, we propose a new multilinear supervised projection method, called Multilinear Enhanced Fisher Discriminant Analysis (MEFDA), for multimodal (2D + 3D) face verification. This method improves the generalization capacity by decomposing the discriminant analysis process into a simultaneous diagonalization of the inter-class and intra-class matrices resulting from the deployment of tensor data. In addition, MEFDA uses the discriminant tensor criterion based on an iterative procedure to obtain the optimal projection matrices in each tensor mode. Furthermore, in this method, the original structure and the natural geometry of the data are respected. The proposed algorithm was evaluated on three 3D face databases, FRGC V2.0, CASIA 3D and Bosphorus, in the difficult context of poses, expressions, illuminations and occlusions. Experimental results demonstrate that the proposed method significantly outperforms state-of-the-art methods. In the second part, we propose a new semi-supervised multilinear method for face verification and kinship in uncontrolled environments. We call this new algorithm, Multilinear Side information based Discriminate Analyse (MSIDA) and use it to solve the problem of face-pair matching, which is characterized by weakly labeled data. This new method can naturally represent facial images and extract useful information directly from tensor data rather than matrices or vectors. In the MSIDA method, several inter connected subspaces witch obtained on different tensor modes, so that the subspaces are learned iteratively by unfolding the tensor in the different modes. Using only weak label information, MSIDA projects the face tensor into a new subspace in which the discrimination between classes is improved. The experimental evaluation on four face databases (LFW, Cornel KinFace, UB KinFace and TSKinFace) shows that our approach is significantly surpasses the current state of the art.

Keywords : Face Recognition, Multidimensional Analysis, MEFDA, MSIDA, Multilinear Learning, Tensor Representation.

Résumé

Au cours de ces dernières décennies, la biométrie a fait l'objet d'une grande attention en raison du besoin sans cesse croissant d'authentification d'identité. Parmi diverses modalités biométriques, le visage offre des avantages compétitifs, car l'acquisition de données faciales est naturelle, non intrusive et bien acceptée par les humains. L'utilisation des techniques d'apprentissage de sous-espaces multilinéaires dans la reconnaissance de visage a suscité un vif intérêt par la communauté scientifique pour la réduction de la dimensionnalité et la classification des données multidimensionnelles. Cette thèse est consacrée à la reconnaissance automatique de visage basée sur l'analyse multidimensionnelle. Dans la première partie, nous proposons une nouvelle méthode de projection supervisée multilinéaire, appelée MEFDA (Multilinear Enhanced Fisher Discriminant Analysis), pour la vérification du visage multimodale (2D+3D). Cette méthode améliore la capacité de généralisation en décomposant le processus d'analyse discriminante en une diagonalisation simultanée des matrices de dispersion inter-classes et intra-classes, résultant du déploiement des données tensorielles. En plus, MEFDA utilise le critère de tenseur discriminant basé sur une procédure itérative afin d'obtenir les matrices de projection optimale dans chaque mode du tenseur. En outre, dans cette méthode, la structure originale et la géométrie naturelle des données sont respectées. L'algorithme proposé a été évalué sur trois bases de données de visages 3D, FRGC V2.0, CASIA 3D et Bosphorus, dans le contexte difficile des poses, expressions, illuminations et occlusions. Les résultats expérimentaux démontrent que la méthode proposée surpasse de manière significative les méthodes de l'état de l'art. Dans la deuxième partie, nous proposons une nouvelle méthode multilinéaire semi supervisée pour la vérification de visage et la parenté dans les environnements non contrôlés. Nous appelons le nouvel algorithme MSIDA (Multilinear Side-Information based Discriminant Analysis) et l'utilisons pour résoudre le problème de la correspondance de paires de visages, qui sont caractérisées par des données faiblement étiquetées. Cette nouvelle méthode permet de représenter naturellement des images multidimensionnelles et d'extraire des informations utiles directement à partir de données tensorielles plutôt que de matrices ou de vecteurs. Dans la méthode MSIDA, plusieurs sous-espaces inter-reliés sont obtenus sur différentes modes de tenseurs, de sorte que les sous-espaces sont appris de manière itérative en dépliant le tenseur dans les différentes modes. En utilisant uniquement les informations d'étiquette faibles, MSIDA projette le tenseur de visages dans un nouvel sous-espace dans lequel la discrimination est améliorée. L'évaluation expérimentale sur quatre bases de données de visages (LFW, Cornell KinFace, UB KinFace et TSKinFace) montre que l'approche proposée surpasse de manière significative l'état de l'art actuel.

Mots-clés : Reconnaissance de visage, Analyse multidimensionnelle, MEFDA, MSIDA, Apprentissage multilinéaire, Représentation tensorielle.

Table des matières

Dédicaces	ii
Remerciements	iii
Publications & Communications associées à la thèse	iv
المخلص	v
Abstarct	vi
Résumé	vii
Table des matières	viii
Liste des figures	xii
Liste des tableaux	xiv
Abréviations	xvi
Notations mathématiques	xvii

Chapitre 1 Introduction

1.1 Contexte et motivation	2
1.2 Objectifs et contributions originales	4
1.3 Reconnaissance biométrique	6
1.4 Pourquoi la biométrie faciale ?	7
1.5 La reconnaissance faciale de la parenté	10
1.5.1 Applications de la reconnaissance de la parenté	11
1.5.2 Difficultés de la reconnaissance de visage et la parenté	11
1.5.2.1 Changement d'illumination	12
1.5.2.2 Pose de la tête	13
1.5.2.3 Occultations	13
1.6 Evaluation de la performance d'un système biométrique	14
1.6.1 Mesure des performances pour la reconnaissance du visage	14
1.6.2 Mesure des performances la reconnaissance faciale de la parenté	15
1.7 Articulation de la thèse	15

Chapitre 2 Propriétés de l'algèbre multilinéaire

2.1 Introduction	19
2.2 Bases d'algèbre multilinéaire	19
2.2.1 Notations	19
2.2.2 Définition mathématique	21
2.2.2.1 Vectorisation : transformation tenseur en vecteur	21
2.2.2.2 Déploiement de tenseurs	22
2.2.2.3 Produit intérieur tensoriel et norme de Frobenius	23
2.2.2.4 Produit extérieur	23
2.2.2.5 Produit mode k : multiplication tenseur matrice	24
2.2.2.6 Produit de Kronecker	25
2.2.2.7 Mesure de distance tenseur	26

2.3 Projections multilinéaires	28
2.3.1 Projection vecteur en vecteur	28
2.3.2 Projection tenseur en tenseur	29
2.3.2 Projection tenseur en vecteur	31
2.4 Relations entre les projections multilinéaires	32
2.4.1 Nombre de paramètres à estimer	33
2.5 Apprentissage de sous-espace linéaire	34
2.5.1 Définition du problème	35
2.6 Apprentissage de sous-espace multilinéaire	36
2.6.1 Définition du problème	36
2.6.2 Mesures de dispersion pour les tenseurs	36
2.7 Réduction de la dimensionnalité par l'apprentissage de sous-espace	38
2.8 Conclusion	41

Chapitre 3 Etat de l'art sur l'analyse multilinéaires pour la reconnaissance faciale

3.1 Introduction	43
3.2 Historique d'apprentissage de sous-espaces multilinéaire	43
3.3 Revue sur les algorithmes d'apprentissage de sous-espaces multilinéaires pour la reconnaissance faciale	44
3.3.1 Approches d'apprentissage de sous-espaces multilinéaires non supervisées	46
3.3.1.1 Techniques d'apprentissage de sous-espaces multilinéaires non supervisées par la projection tenseur en tenseur	46
3.3.1.2 Techniques d'apprentissage de sous-espaces multilinéaires non supervisées par la projection tenseur en vecteur	57
3.3.3 Approches d'apprentissage sous-espaces multilinéaires supervisées	58
3.3.3.1 Techniques d'apprentissage de sous-espaces multilinéaires supervisées par la projection tenseur en tenseur	59
3.3.3.2 Techniques d'apprentissage de sous-espaces multilinéaires supervisées par la projection tenseur en vecteur	63
3.4 Conclusion	67

Chapitre 4 MEFDA pour la vérification multimodale du visage 2D+3D robuste

4.1 Introduction	69
4.2 MEFDA (Multilinear Enhanced Fisher Discriminant Analysis)	72
4.3 Aperçu de l'approche de vérification de visage proposé	77
4.3.1 Conception du tenseur basé sur les caractéristiques locales	77
4.3.1.1 Descripteur LPQ (Local Phase Quantization)	78
4.3.2 Réduction de la dimensionnalité et classification	81
4.4 Evaluation expérimentale	82
4.4.1 Base de données de visage	83
4.4.1.1 Base de données FRGC V2.0	83

4.4.1.2 Base de données CASIA 3D	83
4.4.1.3 Base de données Bosphorus	84
4.4.2 Paramètres de réglage	84
4.4.3 Résultats et discussion	85
4.4.3.1 Vérification uni-modale de visage 2D et 3D	85
4.4.3.2 Vérification multimodal de visage 2D + 3D	86
4.4.3.3 MEFDA vis à vis EFM	87
4.4.3.4 Le temps de complexité	89
4.4.3.5 MEFDA vis à vis toutes les approches	89
4.4.3.6 Comparaison avec l'état de l'art	91
4.5 Conclusion	93

Chapitre 5 MSIDA pour la vérification du visage et de la parenté

5.1 Introduction	96
5.2 Etat de l'art	97
5.3 MSIDA (Multilinear Side-Information based Discriminant Analysis)	98
5.3.1 SILD	98
5.3.2 MSIDA	99
5.4 Vérification des paires de visages basés sur MSIDA	103
5.4.1 Extraction de caractéristiques	104
5.4.1.1 Descripteur LPQ (Local Phase Quantization)	104
5.4.1.2 Descripteur BSIF (Binarized Statistical Image Features)	105
5.4.2 Matching	106
5.5 Expériences	107
5.5.1 Base de données de visage	107
5.5.1.1 Base de données LFW (Labeled Faces in the Wild)	107
5.5.1.2 Base de données Cornell KinFace	108
5.5.1.3 Base de données UB KinFace	108
5.5.1.4 Base de données TSKinFace	109
5.5.2 Paramètres de réglage	110
5.5.3 Résultats et discussion	110
5.5.4 Effet de la quantité de paires d'apprentissage	113
5.5.5 MSIDA faiblement supervisé contre LDA et MDA supervisé	115
5.5.6 Temps de complexité	116
5.5.7 Comparaison avec l'état de l'art	117
5.5.7.1 Vérification de visage dans les environnements non contrôlés	117
5.5.7.2 Vérification de parenté dans les environnements non contrôlés	119
5.6 Conclusion	120
6.1 Conclusion générale	123
6.2 Perspectives	125
Bibliographie	128

Liste des figures

Chapitre 1 Introduction

Figure 1.1 Exemple de modalités biométriques. (Les images sont recueillies à partir d'Internet).....	7
Figure 1.2 Échantillons de visage humain capturés selon différentes techniques.....	9
Figure 1.3 Exemples d'images de visage dans la nature difficile.....	12

Chapitre 2 Propriétés de l'algèbre multilinéaire

Figure 2.1 Exemple d'une représentation tensorielle : (a) deuxième ordre, (b) troisième ordre, (c) quatrième ordre (Séquence vidéo couleur).....	20
Figure 2.2 Illustration visuelle des vecteurs en mode k: (a) $\mathbf{A} \in \mathbb{R}^{8 \times 6 \times 4}$, (b) vecteurs mode 1, (c) vecteurs mode 2, (d) vecteurs mode 3.....	21
Figure 2.3 Déploiement d'un tenseur d'ordre 3 dans les différents modes.....	22
Figure 2.4 Illustration visuelle d'un produit mode 1 du tenseur X du troisième ordre avec une matrice G.....	25
Figure 2.5 Illustration de la projection de vecteur en vecteur.....	29
Figure 2.6 Façon alternative de projeter un tenseur dans un espace de petite dimension.....	29
Figure 2.8 Illustration de la projection de tenseur en tenseur: (a) projection d'un tenseur dans tous les modes (b) projection d'un tenseur dans un seul mode.....	30
Figure 2.9 Illustration de la projection de tenseur en vecteur par projection élémentaire multilinéaire: (a) projection de tenseur en vecteur et (b) projection élémentaire multilinéaire.....	32

Chapitre 3 Etat de l'art sur l'analyse multilinéaires pour la reconnaissance faciale

Figure 3.1 Taxonomie des algorithmes d'apprentissage de sous-espaces multilinéaires non supervisé.....	45
Figure 3.2 Vue schématique de la différence essentielle entre GPCA et PCA.....	47
Figure 3.3 Regroupement d'un sous-ensemble de la base de données de visages Yale B sous des conditions d'illumination pour les sujets 2, 5 et 8. (a) Données d'image projetées sur les trois composants principaux. (b) Résultats de regroupement par la méthode GPCA.....	48
Figure 3.4 Illustration de la méthode Rank-R Approximations of Tensors vis à vis le GLRAM.....	51
Figure 3.5 Réduction de la dimensionnalité et de la reconstruction des objets pour les matrices et les tenseurs d'ordre supérieur. (1) indique une réduction de la dimensionnalité, (2) reconstruction d'objet par l'algorithme CSA.....	52
Figure 3.6 Schéma fonctionnel du système de reconnaissance d'objets tensoriel basé sur la méthode MPCA.....	54
Figure 3.7 Illustration d'échantillons à deux classes dans un espace à deux dimensions et les directions de projection avec MPCA, TLPP, et MSDA.....	55
Figure 3.8 Résultats de la visualisation de différentes méthodes sur la base de données OR.....	56

Figure 3.9 Taxonomie des algorithmes d'apprentissage de sous-espaces multilinéaires supervisés.	58
Figure 3.10 L'évolution du critère objectif de la méthode DATER au fil des itérations.....	61
Figure 3.11 Variance inter-classe et variance intra-classe avant et après l'utilisation MDA.	63
Figure 3.12 Reconnaissance des objets (visages) tenseurs basée sur UMLDA	65
Figure 3.13 Illustration des trois premières projections multilinéaires élémentaires par UMLDA de (a) la base de données PIE, (b) la base de données FERET.	66

Chapitre 4 MEFDA pour la Vérification multimodale du visage 2D+3D robuste

Figure 4.1 Système proposé pour la vérification multimodale de visage 2D+3D.	78
Figure 4.2 Organigramme de l'ensemble des étapes nécessaires du descripteur LPQ.	80
Figure 4.3 Extraction des caractéristiques et conception du tenseur.....	82
Figure 4.4 Exemples d'images de visage de la base de données FRGC V2.0.....	83
Figure 4.5 Exemples d'images de visage de la base de données CASIA 3D.	84
Figure 4.6 Exemples d'images de visage de la base de données Bosphorus.	84
Figure 4.7 Taux de vérification de EFM et MEFDA pour 2D, 3D et 2D + 3D sur (a) FRGC V2.0 (b) CASIA 3D (c) Bosphorus, respectivement.	88
Figure 4.8 Courbe ROC multimodale pour différentes transformations de sous-espaces sur la base de données CASIA 3D.....	90
Figure 4.9 Courbe ROC multimodale pour différentes transformations de sous-espaces sur la base de données Bosphorus.	90
Figure 4.10 Courbes ROC multimodale pour différentes transformations de sous-espaces sur la base de données FRGC V2.0.	91

Chapitre 5 MSIDA pour la vérification du visage et de la parenté

Figure 5.1 Schéma fonctionnel du système de matching des paires de visage proposé.....	103
Figure 5.2 Représentation d'image du visage avec le descripteur LPQ sous différentes tailles de fenêtre m.....	105
Figure 5.3 Les 13 images naturelles utilisées pour l'apprentissage des filtres dans le descripteur BSIF	106
Figure 5.4 Représentation d'image du visage avec le descripteur BSIF sous différentes tailles du filtre l.....	106
Figure 5.5 Exemples d'images de la base de données LFW. Gauche : paires positives et droite : paires négatives.....	108
Figure 5.6 Exemples d'images de la de données Cornell KinFace. Gauche: paires positives et droite: paires négatives.	108
Figure 5.7 Exemples d'images de la de données UB KinFace. Gauche : paires positives et droite : paires négatives.	109
Figure 5.8 Exemples d'images de la de données TSKinFace. Gauche : paires positives et droite : paires négatives.	109
Figure 5.9 Précision moyenne de SILD et MSIDA avec un nombre différent de paires sur les bases de données (a) LFW (b) Cornell KinFace (c) UB KinFace (d) TSKinFace, respectivement.	114

Figure 5.10 Courbes ROC des différentes méthodes selon les protocoles restreint et non restreint sur la base de données LFW.	116
Figure 5.11 Courbe ROC de MSIDA et état de l'art sur la base de données LFW sous le protocole le restreint.	118

Liste des tableaux

Chapitre 1 Introduction

Tableau 1.1 Brève comparaison des traits biométriques, Haut, Moyen et Bas sont désignés par H, M et B, respectivement	8
--	---

Chapitre 2 Propriétés d’algèbre multilinéaire

Tableau 2.2 Exemples de modélisations tensorielles.	21
Tableau 2.2 Nombre de paramètres à estimer selon trois projections multilinéaires.	34

Chapitre 4 MEFDA pour la vérification multimodale du visage 2D+3D robuste

Tableau 4.1 Taux de vérification (%) à 0,001 TFA uni-modale 2D, 3D et multimodal 2D + 3D sur la base FRGC V2.0.	86
Tableau 4.2 Performances de la vérification (%) uni-modale 2D, 3D et multimodales 2D + 3D sur la base CASIA 3D.	87
Tableau 4.3 Taux de vérification (%) à 0,001 TFA uni-modale 2D, 3D de visage et multimodal 2D + 3D sur la base Bosphorus.	87
Tableau 4.4 Temps de complexité en secondes d’EFM et MEFDA.	89
Tableau 4.5 Comparaison de taux de vérification (%) avec l’état de l’art sur la base de données FRGC V2.0.	92
Tableau 4.6 Comparaison de taux de vérification (%) avec l’état de l’art sur la base de données CASIA 3D.	93
Tableau 4.7 Comparaison de taux de vérification (%) avec l’état de l’art sur la base de données Bosphorus.	93

Chapitre 5 MSIDA pour la vérification du visage et de la parenté

Tableau 5.1 Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données LFW.	111
Tableau 5.2 Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données Cornell KinFace.	111
Tableau 5.3 Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données UB KinFace.	112
Tableau 5.4 Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données TSKinFace.	112
Tableau 5.5 Précision de la vérification (%) des différentes méthodes selon les protocoles restreint et non restreint sur la base de données LFW.	115
Tableau 5.6 Temps de complexité en secondes de SILD et MSIDA sur les quatre bases de données.	117
Tableau 5.7 Comparaison la précision de vérification de MSIDA avec l’état de l’art sur base de données LFW selon le protocole restreint.	118

Tableau 5.8 Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données Cornell KinFace.	119
Tableau 5.9 Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données UB KinFace.	120
Tableau 5.10 Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données TSKinFace.	120

Abréviations

- CMC:** Cumulative Match Characteristic.
- R1RR:** Rank-1 Recognition Rate.
- TFR:** Taux de Faux Rejets.
- TFA:** Taux de Fausses Acceptations.
- TEE:** Taux d'Égale Erreur.
- ROC:** Receiver Operating Characteristic.
- TV:** Taux de Vérification.
- BSIF:** Binarized Statistical Image Features.
- LPQ:** Local Phase Quantization.
- PCA:** Principal Component Analysis.
- LDA:** Linear Discriminant Analysis.
- LPP:** Locality Preserving Projection.
- NPP:** Neighborhood Preserving Projection.
- CCA:** Canonical Correlation Analysis.
- ICA:** Independent Component Analysis.
- SILD:** Side-Information based Discriminant Analysis.
- SIEDA:** Side-Information based Exponential Discriminant Analysis.
- 2DPCA:** Two-Dimensional Principal Component Analysis.
- GPCA:** Generalized Principal Component Analysis.
- BD-PCA:** Bidirectional Principal Component Analysis.
- GLRAM:** Generalized Low Rank Approximation of Matrices.
- CSA:** Concurrent Subspaces Analysis.
- MPCA:** Multilinear principal Component Analysis.
- TLPP:** Tensor Locality Preserving Projection.
- MSDA:** Multilinear Spatial Discriminant Analysis.
- TROD:** Tensor Rank-One Decomposition.
- 2DLDA:** Two-Dimensional Linear Discriminant Analysis.
- DATER:** Discriminant Analysis with Tensor Representation.
- GTDA:** General Tensor Discriminant Analysis.
- MDA:** Multilinear Discriminant Analysis.
- TR1DA:** Tensor Rank-One Discriminant Analysis.

UMDLA: Uncorrelated Multilinear Discriminant Analysis.

LMKMML: Localized Multi-Kernel Multi-Metric Learning.

MEFDA: Multilinear Enhanced Fisher Discriminant Analysis.

MSIDA: Multilinear Side-Information based Discriminant Analysis.

Notations mathématiques

$a \in \mathbb{R}$: Scalaire.

$\mathbf{a} \in \mathbb{R}^{I_1}$: Vecteur de dimension I_1 .

$A \in \mathbb{R}^{I_1 \times I_2}$: Matrice de dimension $I_1 \times I_2$.

$\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$: Tenseur d'ordre N de dimension $I_1 \times I_2 \times \dots \times I_N$.

I_K : Dimension de l'espace vectoriel k -modal.

$a_{i_1 i_2 \dots i_N}$: Élément d'indice $(i_1 i_2 \dots i_N)$ du tenseur \mathbf{A} .

$A^{(k)}$: Matrice dépliant dans le mode k du tenseur \mathbf{A} .

\circ : Produit extérieur entre plusieurs vecteurs n -modaux.

\otimes : Produit tensoriel de Kronecker.

$\text{tr}(A)$: Trace de la matrice A .

$\|\mathbf{A}\|_F$: Norme de Frobenius du tenseur \mathbf{A} .

$\text{vec}(\mathbf{A})$: Représentation vectorielle du tenseur \mathbf{A} .

$\bar{\mathbf{A}}$: Tenseur moyen de toutes les échantillons d'apprentissage.

$U^{(k)}$: Matrice de projection dans le mode k .

S_b : Matrice de dispersion inter-classes.

S_w : Matrice de dispersion intra-classes.

N_b : Nouvelle matrice de dispersion inter-classes.

C : Nombre de classes dans la base de données facial.

Λ : Matrice des vecteurs propres.

λ : Valeur propre.

$\psi_{\mathbf{A}}$: Dispersion totale des tenseurs d'apprentissage.

ε : Petit nombre constant.

itr : Nombre maximum d'itérations dans les solutions multilinéaires.

F_{LPQ} : L'étiquette d'image du descripteur LPQ.

W_i : Filtre de descripteur BSI.

$\cos(\mathbf{z}_{t_1}, \mathbf{z}_{t_2})$: Similarité de cosinus entre deux vecteurs \mathbf{z}_{t_1} et \mathbf{z}_{t_2} .

Chapitre 1

Introduction

1.1 Contexte et motivation	2
1.2 Objectifs et contributions originales	4
1.3 Reconnaissance biométrique	6
1.4 Pourquoi la biométrie faciale ?.....	7
1.4.1 La reconnaissance faciale de la parenté.....	10
1.4.2 Applications de la reconnaissance de la parenté	11
1.4.3 Difficultés de la reconnaissance de visage et la parenté.....	11
1.4.3.1 Changement d'illumination	12
1.4.3.2 Pose de la tête.....	13
1.4.3.3 Occultations	13
1.4.4 Evaluation de la performance d'un système biométrique.....	14
1.4.4.1 Mesure des performances pour la reconnaissance du visage	14
1.4.4.2 Mesure des performances la reconnaissance faciale de la parenté	15
1.5 Articulation de la thèse.....	15

1.1 Contexte et motivation

La reconnaissance automatique des gens en fonction de leurs images faciales numériques est considérée comme un sujet de recherche très important dans le domaine de la vision par ordinateur et la biométrie. Les images faciales contiennent de nombreuses informations utiles sur la personne à laquelle ils appartiennent, notamment l'identité, sexe, ethnicité, âge, et expression émotionnelle, etc. Ainsi, l'image du visage est importante pour la plupart des systèmes biométriques. Pour cette raison, la reconnaissance du visage, est un sujet de recherche essentiel pour la communauté scientifique en intelligence artificielle depuis plus de vingt ans, les premières études théoriques sur la reconnaissance faciale remontent au moins au début des années 1970. Le passage vers des produits commerciaux n'a reçu une impulsion décisive qu'à partir des années 1994-1996 en bonne partie grâce à la mise en œuvre d'un programme d'évaluation international FERET (Face Recognition Technology), organisé par le ministère de la défense américaine (Department of Defense, DoD) [1]. Le vif intérêt du marché potentiellement gigantesque a fait éclore un grand nombre d'algorithmes de reconnaissance de visages, Le DoD décida alors de lancer des programmes d'évaluation 2000, 2002 et en 2006 dont l'objectif était d'évaluer les performances des produits commerciaux.

Un système biométrique facial se décompose en de deux phases, la phase d'apprentissage (off-line) et la phase de test (on-line). La phase d'apprentissage sera réalisée une seule fois dans laquelle l'enrôlement des images faciales des différentes personnes est utilisé afin d'extraire la signature biométrique de chaque individu. Pendant la phase de test, les nouvelles données sont comparées avec les données d'entraînement qui sont apprises dans la phase d'apprentissage afin de prendre une décision d'accepter ou rejeter le candidat. Les étapes effectuées dans ces deux phases dans un système de reconnaissance du visage sont subdivisées en trois modules principaux [2] : la détection des visages, l'extraction de caractéristiques et la reconnaissance (classification).

Généralement, les applications en traitement d'images comportent des données multidimensionnelles relativement massives. Ces données naturelles sont une sélection composite de multiples facteurs et modes liés à la structure de la scène, illumination, expressions faciales, poses de tête. Les objets multidimensionnels sont officiellement nommés données tensorielles. L'expression (tenseur) a deux significations différentes dans les domaines de la physique et des mathématiques. Dans cette thèse, le mot tenseur se réfère à la même notation utilisée en mathématiques multilinéaires. Ainsi, un tenseur peut être défini

mathématiquement comme la génération de vecteurs et d'une matrice. Comme il est bien connu, lorsque les données consistent en une dimension, elles sont définies comme un vecteur et, lorsqu'elles sont composées de deux dimensions, elles sont définies comme une matrice et lorsqu'elles sont composées de plus de deux dimensions, elles sont connues sous le nom du tenseur [3].

Naturellement, des objets différents ont des structures spécifiques qui se présentent sous la forme de tenseurs d'ordre deux, trois ou même supérieurs, une image de visage de niveau de gris se présente comme un objet tenseur bidimensionnel tandis qu'une image couleur ainsi que les données de séquence des images grises (niveau de gris) d'une vidéo se présentent sous la forme d'objets tenseurs du troisième ordre. Ainsi, les algorithmes nécessitent d'extraire des informations utiles à partir de ces données multidimensionnelles. C'est un problème difficile en raison de la variabilité et de la complexité de la distribution des motifs faciaux. Les approches de reconnaissance faciale qui agissent directement sur un tel espace de grande dimension souffrent généralement du problème appelé malédiction de la dimensionnalité. L'essence de ce problème est que les données de grandes dimensions sont difficiles à manipuler directement en raison de l'ajout de caractéristiques supplémentaires, cela peut augmenter le bruit et donc augmenter l'erreur résultante, diminuant ainsi le taux de précision, en plus d'être très coûteux en calcul [4]. L'apprentissage de sous espace ou la réduction de dimensionnalité vise à mapper l'ensemble de données de grande dimension vers un ensemble de dimension inférieure tout en conservant les informations les plus distinctives et en préservant les propriétés particulières telles que les informations géométriques de voisinage globales et locales [5].

Les méthodes traditionnelles d'apprentissage de sous espace linéaire sont basées sur des algorithmes unidimensionnels. Diverses méthodes ont été utilisées pour la reconnaissance faciale au cours des deux dernières décennies, telles que PCA (Principal Component analysis) [6], LDA (Linear Discriminant Analysis) [7], LPP (Local Preserving Projection) [8], NPP (Neighborhood Preserving Projection) [9]. Alors que PCA recherche les directions de représentation les plus efficaces, LDA cherche la direction de discrimination efficace et LLP se base sur la recherche des approximations linéaires optimales [10]. D'autre part, l'algorithme NPP utilise des pondérations basées sur les données en résolvant un problème des moindres carrés afin de refléter la géométrie intrinsèque des voisinages locaux. L'application des techniques d'apprentissage de sous-espace aux données faciales tensorielles nécessiterait d'abord de remodeler ces données en un vecteur 1-D. Cette vectorisation ignore la structure sous-jacente des données en raison de la destruction de la structure naturelle et des corrélations

internes existantes dans les données multilinéaires originales, en plus perdre la représentation la plus utile des données faciales tensorielles. De plus, cette transformation de données nécessite généralement plus de temps de calcul, ce qui pose souvent le problème de la dimensionnalité, et entraîne une réduction du taux de reconnaissance.

D'autre part, les méthodes d'apprentissage de sous espaces multilinéaires fonctionnent sur des objets tensoriels naturels et extraient des caractéristiques directement à partir de la représentation tensorielle. C'est donc une option prometteuse pour le traitement des objets tenseurs faciaux. Au cours des dernières années, l'algèbre multidimensionnelle, en tant qu'algèbre des tenseurs d'ordre supérieur, a fait l'objet d'une attention particulière, de nombreux chercheurs ayant commencé à représenter les données sous leur forme naturelle [11-22]. Ces méthodes extraient immédiatement les caractéristiques de données faciales tensorielles sans avoir besoin de remodeler les données d'objet en un seul vecteur. Par conséquent, cette représentation a été appliquée à l'analyse de la structure multifactorielle d'images [23]. Cela a inspiré le développement récent d'algorithmes d'apprentissage de sous-espaces multilinéaires, telles que MPCA (Multilinear Principal Component Analysis) [24], TLPP (Tensor Locality Preserving Projection) [21], TROD (Tensor Rank-One Decomposition) [25], MDA (Multilinear Discriminant Analysis) [12]. Ces méthodes extraient directement les caractéristiques utiles et les plus compactes de données faciales tensorielles au lieu de les remodeler en un seul vecteur.

A travers les différents chapitres, nous mettons en avant l'intérêt de l'utilisation d'algorithmes basées sur l'analyse multidimensionnelle à l'aide du formalisme tensoriel dans la reconnaissance de visage.

1.2 Objectifs et contributions originales

La thèse porte principalement sur le développement, l'implémentation et l'évaluation d'un système automatique et efficace de reconnaissance de visage basé sur les techniques d'apprentissage de sous espaces multilinéaires dans les environnements non contrôlés dans lesquels les variations : d'illuminations, d'expressions et de poses sont très différentes entre l'ensemble d'apprentissage et l'ensemble de test.

À la lumière de la discussion précédente, cette thèse propose les contributions de recherche principales comme suit :

- Etude de l'état de l'art des différentes techniques de reconnaissance de visage basées sur l'analyse multidimensionnelle.
- Etude et conception d'un système efficace de vérification de visage multimodale (2D+3D) aux variations d'expression, d'illumination et de pose, basées sur la représentation des tenseurs d'ordre élevé, en utilisant les caractéristiques locales extraites de l'image intensité et l'image profondeur.
- Proposition d'une nouvelle méthode de projection multilinéaire supervisée, appelée MEFDA (Multilinear Enhanced Fisher Discriminant Analysis). Cette approche représente naturellement les objets multidimensionnels en tant que tenseurs d'ordre supérieur et mappe un tenseur de grande dimension à un espace de données de faible dimension en appliquant une réduction dans les modes du tenseur. Cet algorithme améliore la capacité de généralisation, en décomposant le processus d'analyse discriminante dans une diagonalisation simultanée des matrices de dispersion inter-classes et intra-classes, résultant du déploiement des données tensorielles. L'un des avantages de MEFDA est qu'il peut éviter l'inconvénient d'une grande dimensionnalité, dans laquelle le coût de calcul est réduit. Puisque les données encodées par la représentation tensorielle, MEFDA respecte la structure originale et la géométrie naturelle des données faciales. Par conséquent, la corrélation naturelle des données multidimensionnelles de visage peut être bien exploitée.
- Un certain nombre d'expériences ont été réalisées sur trois bases de données faciales universelles (FRGC v2.0, CASIA 3D et Bosphorus), afin d'évaluer les performances d'algorithmes proposés MEFDA, et de prouver la supériorité de MEFDA par rapport les méthodes de l'état de l'art existantes telles que : PCA, LDA, EFM, MPCA et MDA.
- Proposition d'une nouvelle méthode multilinéaire faiblement supervisée (semi supervisée) pour la vérification faciale et la parenté appelée MSIDA (Multi-Linear Side-Information based Discriminant Analysis). Cet algorithme adresse la projection et la classification des sous-espaces multilinéaires faiblement supervisées. En utilisant uniquement les informations d'étiquettes faibles, MSIDA projette le tenseur de visages dans un nouvel sous-espace dans lequel la discrimination est améliorée et la dimension de chaque mode du tenseur est réduite simultanément. Aussi bien que, MSIDA représente naturellement les objets multidimensionnels.

- Etude et conception d'un système robuste de vérification de la parenté basée sur l'extraction de caractéristiques multi-échelles qui sont projetées dans le sous-espace MSIDA.
- Proposition d'une nouvelle conception tensorielle des données de visages. Cette représentation combine différents descripteurs locaux, extraits à différentes échelles, permettant une meilleure discrimination. Ainsi, la représentation du tenseur proposée est considérée comme une nouvelle stratégie de fusion des descripteurs faciaux.
- Des expériences approfondies ont été effectuées sur quatre bases de données difficiles (LFW, Cornell KinFace, UB KinFace et TSkinFace), afin d'examiner les performances de la méthode proposée MSIDA.

1.3 Reconnaissance biométrique

La biométrie est la reconnaissance automatique des personnes à partir de leurs caractéristiques physiologiques (empreintes digitales, visage, iris, contour de la main,..... etc.), ou comportementales (signature, démarche,.....etc.). Ces caractéristiques s'appellent modalités biométriques. La figure 1.1 montre les différents types des modalités biométriques. Au cours des dernières années, la reconnaissance biométrique a été une des solutions les plus appropriées pour les applications nécessitant une haute sécurité telle que l'accès aux aéroports, aux centrales nucléaires, contrôle des frontières, banques,.... etc.[26-29]. Cependant, l'utilisation de la biométrie pour l'authentification personnelle devient pratique et beaucoup plus précise que les méthodes traditionnelles comme l'utilisation des mots de passe, de la carte d'identité, du code PIN ou des cartes à puce qui peuvent être volées ou falsifiées. En effet, pour qu'un système biométrique puisse fonctionner en environnement réel, les caractéristiques physiologiques doivent satisfaire sept critères identifiés par Jain et al [30] comme suit

1. **Universalité** : chaque personne doit avoir la caractéristique choisie.
2. **Unicité** : la caractéristique doit être suffisamment unique pour distinguer une personne d'une autre.
3. **Permanence** : la caractéristique doit être invariante avec un individu sur une longue période.
4. **Mesurabilité** : la caractéristique doit être facile à capturer et adaptée à la mesure.
5. **Acceptabilité** : la caractéristique doit être acceptée par les personnes et ne pas les ennuyer.

6. **Performance** : la précision, la vitesse de traitement et la robustesse du système biométrique impliquant la caractéristique sont suffisamment satisfaisantes pour les demandes d'authentification.
7. **Contournement** : la caractéristique doit être difficile à imiter avec un artéfact ou un substitut.

Cependant, dans la pratique, aucune modalité biométrique ne répondra à toutes ces exigences [30]. Par conséquent, une sélection de la modalité biométrie que doit être choisie pour toute application possible en fonction d'indicateurs tels que la disponibilité des capteurs et des appareils, le temps de calcul requis, la fiabilité, le coût, et la consommation d'énergie.



Figure 1.1: Exemple de modalités biométriques. (les images sont recueillies à partir d'Internet).

1.4 Pourquoi la biométrie faciale ?

Parmi toutes les modalités biométriques utilisés pour la reconnaissance des personnes, l'analyse du visage a récemment fait l'objet de beaucoup d'attention en raison des énormes développements dans le domaine du traitement d'images et de l'apprentissage automatique. La reconnaissance de visages est la technique la plus commune et populaire. Elle reste la plus acceptable puisqu'elle correspond à ce que les humains utilisent dans l'interaction visuelle, au-delà de son intérêt scientifique, comparé à d'autres éléments biométriques tels que les empreintes digitales et l'iris, le visage offre de nombreux avantages incomparables pour une grande variété d'applications potentielles dans le commerce et l'application de la loi. Le tableau 1.1 présente une comparaison du visage et d'autres modalités biométriques, basée sur les sept critères précédents [30].

Tableau 1.1: Brève comparaison des traits biométriques, Haut, Moyen et Bas sont désignés par H, M et B, respectivement [31].

Modalité biométrique	Visage	Empreinte digitale	Iris	Voix	ADN	Empreinte palmaire	Oreille	démarche	Rétine	Signature
Unicité	M	H	H	B	H	H	M	B	H	B
Universalité	H	M	H	M	H	M	M	M	H	B
Permanence	M	M	H	B	H	H	H	B	M	B
Mesurabilité	H	M	M	M	B	M	M	H	B	H
Acceptabilité	H	M	B	H	B	M	H	H	B	H
Contournement	H	M	B	H	B	M	M	M	B	H
Performance	H	H	H	B	H	H	M	B	H	B

À partir de ce tableau, nous pouvons en déduire que le visage est supérieur aux autres données biométriques pour les raisons suivantes :

Processus non intrusif : au lieu de demander aux utilisateurs de placer leur main ou leurs doigts sur un lecteur (processus inacceptable dans certaines cultures et source de transfert de la maladie) ou de positionner leurs yeux avec précision devant un scanner, les systèmes de reconnaissance de visage prennent discrètement des photos de visages de personnes à distance.

Aucune intrusion ou délai nécessaire et dans la plupart des cas, les utilisateurs ignorent totalement le processus de capture. Ils ne sentent pas que leur vie privée a été envahie ou «sous surveillance». De plus, identifier une personne sur la base de son visage est l'un des types de comportement humain les plus anciens et les plus élémentaires, ce qui le rend naturellement accepté par le public.

Facilité de mise en œuvre : contrairement à la plupart des modalités biométriques qui nécessitent du matériel professionnel lors de leur mise en œuvre (lecteur numérique et scanner d'empreintes digitales, de paume, d'iris et de rétine, par exemple), les données de visage peuvent être facilement capturées via des appareils photo numériques, des appareils photo sur PC ou même l'utilisation répandue de smartphones.

Diverses techniques : contrairement à d'autres modalités biométriques qui sont normalement unimodaux (principalement des images couleur / niveaux de gris), les données de visage

peuvent être capturées et stockées selon diverses techniques. Différentes techniques sont exploitées dans différents scénarios de reconnaissance de visage en fonction de leurs propres caractéristiques. Les images couleur/ niveaux de gris 2D suffisent pour les tâches de reconnaissance normales, les images en profondeur et les numérisations 3D sont plus robustes contre les variations d'éclairage, les images croquis de visages sont largement utilisés dans les enquêtes sur les crimes graves commis par la police. Plus précisément, la collaboration entre les images 2D et les modèles 3D améliore considérablement les performances de reconnaissance des visages. Plusieurs modalités couramment utilisées sont illustrées à la figure 1.2.

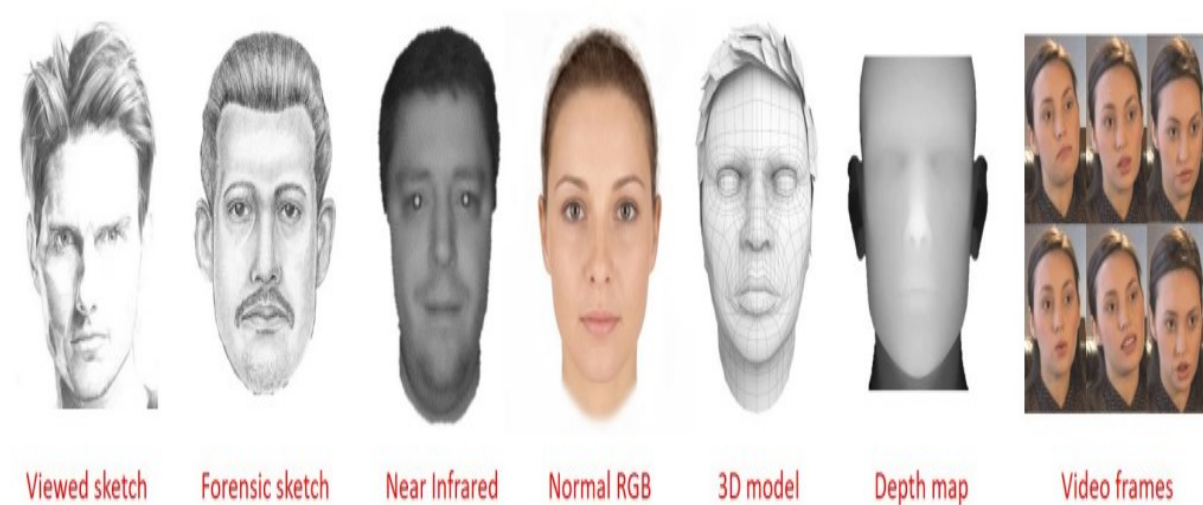


Figure 1.2: Echantillons de visage humain capturés selon différentes techniques [31].

Perception biologique : c'est peut-être la plus importante. Les humains qui identifient d'autres personnes par leur visage et leur voix sont donc susceptibles d'être à l'aise avec les systèmes qui utilisent le visage et la reconnaissance vocale [32].

Amélioration de la performance : jusqu'à la première décennie de ce siècle, les performances d'identification basées sur le visage étaient relativement médiocres par rapport aux performances basées sur d'autres traits forts tels que l'iris et la rétine [26]. La principale raison réside dans la capacité limitée de distinguer une personne dans un environnement incontrôlé, car les représentations faciales peuvent être plus sensibles aux variations d'éclairage, pose, expression, etc. Cependant, au cours des dernières années, la reconnaissance des visages a considérablement progressé avec le développement de méthodes de la reconnaissance faciale.

La reconnaissance des visages ne montre également aucun point faible sous aucun aspect, comme le montre le tableau 1.1, ce qui en fait un candidat stablement et raisonnablement accepté comme candidat principal pour tous les traits biométriques. De nos jours, la technologie de reconnaissance des visages fait de plus en plus partie de la vie quotidienne des utilisateurs, sous la forme d'applications pertinentes, notamment le contrôle d'accès, le suivi des suspects, la vidéosurveillance et les interactions homme-machine.

1.5 La reconnaissance faciale de la parenté

Avant d'aborder la nature du problème et les solutions, nous proposons des termes de liens de parenté différents. En général, la parenté peut indiquer une similitude, une familiarité ou une proximité entre des entités sur la base de certains traits ou de toutes les caractéristiques de base. La parenté peut être formulée ou étiquetée comme une relation. Dans le contexte anthropologique, la parenté renvoie aux réseaux sociaux des relations entre les personnes qui forment une partie importante de la vie des humains dans la plupart des sociétés. Néanmoins, dans d'autres contextes, la parenté peut avoir des significations différentes. En biologie, la parenté se réfère généralement au degré de parenté génétique ou coefficient de relation entre membres individuels de la même espèce [33, 34].

La reconnaissance automatique faciale de la parenté est un défi intrinsèque et pour cette raison, un sujet très intéressant à étudier. Lors de l'analyse des images faciales de deux individus, l'ordinateur doit indiquer si ces sujets sont liés ou non. En outre, différents degrés de la parenté pourraient être détectés, en tant que : grands parents-enfant, parents-enfant, mère-frères et sœurs, etc. Dans ce cas, plusieurs complications apparaissent en raison de variations de genre, d'âge et de plus petite similitude entre parents éloignés. Cela caractérise le domaine de la reconnaissance automatique de la parenté comme un problème beaucoup plus large et difficile que l'identification et la vérification de visage traditionnelle. Dans la reconnaissance automatique de parenté à l'aide de techniques de vision par ordinateur, la machine est capable de distinguer les parents de personnes non apparentées et de déterminer le degré de parenté basée sur une inspection de leurs images. En d'autres termes, il s'agit de former une machine à reconnaître la relation de sang entre une paire de visages de parenté et de non parenté (vérification) à partir de traits extraits d'images faciales et de déterminer le type ou le degré exact de cette relation (reconnaissance).

1.5.1 Applications de la reconnaissance de la parenté

La reconnaissance de la parenté a un impact significatif sur les applications réelles et de nombreux domaines. Nous récapitulons les domaines possibles d'application comme suit :

- organiser des collections d'images et résoudre des identités dans des albums photos
- rechercher des parents dans des bases de données publiques
- renforcement des capacités de vérification automatique des visages
- marquage automatique d'un grand nombre d'images disponibles en ligne
- déterminer les parents d'une victime ou d'un suspect par les organismes d'application de la loi.
- Construire un arbre de relations dynamiques à partir d'une organisation de photos et d'albums familiaux.

Cependant, la recherche sur ce sujet reste complexe et difficile. Pour atteindre l'objectif de la vérification de la parenté, l'analyse de la parenté à travers les visages humains doit être menée de façon exhaustive. Les motivations et les applications susmentionnées attirent un nombre important de chercheurs pour présenter leurs contributions dans ce domaine. Évidemment, la recherche dans ce domaine est toujours active.

1.5.2 Difficultés de la reconnaissance de visage et la parenté

Alors que dans la réalité, le visage humain est un sujet à une grande variété de changements intrinsèques et extrinsèques, les systèmes d'analyse de visage sont formés à un nombre limité d'échantillons de visage généralement capturés dans des conditions contrôlées. La figure 1.3 illustre certaines images de visage difficiles capturées dans la nature (conditions incontrôlées). Il est pratiquement impossible de couvrir tous les changements faciaux pendant la phase de l'apprentissage, ce qui empêche les systèmes d'analyse faciale de gérer les visages invisibles avec de nouveaux changements. De plus, des études de littérature [35] ont démontré que des variations en termes (changement d'illumination, pose de la tête, occultations... etc.) dans différentes images de visage d'une même personne peuvent être plus importantes que les variations de visages de personnes différentes. Par conséquent, les performances d'analyse de visage se dégradent remarquablement dans des environnements défavorables. Cette section passe en revue les principaux problèmes qui entravent l'analyse des visages et fait référence aux principales solutions proposées dans la littérature.



Figure 1.3: Exemples d'images de visage en milieux non contrôlés.

1.5.2.1 Changement d'illumination

L'intensité et la direction d'illumination lors de la prise de vue influent toutes deux énormément sur l'apparence du visage dans l'image. Dans la plupart des applications réelles, des changements dans les conditions d'illumination sont néanmoins inévitables, notamment lorsque les vues sont collectées à des dates déferentes. Etant donné qu'un visage humain est un objet intrinsèquement 3D, des changements d'illumination peuvent faire apparaître sur le visage des ombres accentuant ou, au contraire, masquant certaines caractéristiques faciales. Parmi les solutions proposées dans la littérature au problème de l'éclairage, il y a le développement de descripteurs de visage robustes au changement d'illumination. Cependant, l'étude de [36], qui impliquait plusieurs représentations d'image relativement insensibles à l'éclairage sous des changements d'illumination a démontré qu'aucune méthode capablement suffisante pour résoudre le problème.

Certaines recherches ont recours à d'autres technologies de détection moins sujettes au changement de luminosité que les images d'intensité. Par conséquent, des capteurs 3D ont été utilisés pour capturer les images de la plage faciale décrivant la profondeur des objets de la scène. Une alternative aux capteurs 3D consiste à reconstruire le visage à partir d'images 2D à l'aide de techniques de vision par ordinateur et à appliquer un éclairage synthétique au modèle

de visage 3D. Des capteurs (Near Infrared) NIR ont également été étudiés pour résoudre le problème d'illumination du visage.

1.5.2.2 Pose de la tête

Le changement de l'angle d'inclinaison de la tête engendre de nombreux changements d'apparence dans l'image collectée. Les rotations en profondeur engendrent deux types de difficultés. Tout d'abord, elles amènent des différences de profondeur qui projetées sur le plan 2D de l'image, résultent en des déformations (étirement de certaines parties du visage et compactage d'autres régions). Secondement, elles peuvent à l'occultation de certaines parties du visage (par exemple, dans une vue de trois quarts, une partie du visage est cachée).

Les visages 3D, collectés par des scanners 3D ou reconstitués à partir d'images 2D, sont utiles pour traiter les changements des poses de tête. De nombreuses approches pour l'estimation de la pose de la tête ont été proposées dans la littérature [37]. Une fois la pose estimée, la tête peut être tournée dans une position normalisée (très souvent une pose frontale) et le visage est ensuite analysé. On peut également traiter la pose de la tête en construisant un modèle de visage à partir d'images de visage du même individu mais avec des orientations de tête différentes [38] ou en construisant des modèles séparés basés sur la vue pour le même visage [39]. La pose peut également être corrigée en ajustant un modèle 3D sur l'image [40], puis en générant une vue frontale du visage. L'ajustement est réalisé sur la base d'une correspondance de point de repère entre le modèle 3D et l'image du visage. Cette correspondance nécessite une détection automatique des points du visage dans les images 2D.

1.5.2.3 Occultations

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoires tels que des lunettes de soleil, écharpe, chapeau, faux cheveux, etc. Les occultations peuvent donc être intentionnelles ou non. Dans certains cas, il peut s'agir d'une volonté délibérée de contrecarrer la reconnaissance. Les systèmes proposés doivent être non intrusifs, c'est-à-dire qu'on ne peut pas compter sur une coopération du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés.

De nombreux chercheurs ont tenté de gérer de telles situations en proposant des approches robustes pour l'occlusion partielle du visage. Des méthodes de sous-espace ont été utilisées pour projeter le visage dans un nouvel espace et éliminer les parties obstruées. Les descripteurs de visage locaux sont plus robustes que les approches holistiques à l'occlusion partielle du

visage. Les faces sont généralement divisées en petits blocs et chaque bloc est modélisé séparément. Puisque les blocs correspondants sont appariés, seuls les blocs couvrant l'occlusion seront affectés. Certaines autres méthodes [41] ont tenté de reconstruire les parties du visage occluses, tandis que d'autres ([42, 43]) détectent l'occlusion et n'utilisent que des parties non occluses pour l'analyse du visage.

1.6 Evaluation de la performance d'un système biométrique

Les systèmes biométriques doivent être évalués afin d'estimer les différents paramètres. La performance mesure l'efficacité et la fiabilité d'un système biométrique dans un contexte d'utilisation donné. Un ensemble des critères a été proposé dans la littérature pour mesurer les performances d'un système de reconnaissance de visage. Comme indiqué précédemment, il y'a deux scénarios de fonctionnement d'un système biométrique de reconnaissance de visage : identification et vérification, pour chaque mode il y'a différents critères pour évaluer le système.

1.6.1 Mesure des performances pour la reconnaissance du visage

Dans le cas d'une identification, la mesure de performance standard d'un système biométrique utilise les métriques énumérées ci-dessous.

R1RR (Rank-1 Recognition Rate) : il fait référence au pourcentage de modèle de saisie identifié correctement. Le processus de reconnaissance réussit uniquement si le système attribue au modèle d'entrée l'étiquette d'identité correcte.

La courbe CMC (Cumulative Match Characteristic) : il fait référence à une courbe créée en traçant le taux de reconnaissance en fonction des n rangs, n rangs indique que les n premiers modèles inscrits les plus similaires aux modèles de saisie sont considérés comme des candidats. Si l'identité correcte du modèle de saisie est dans le candidat liste, le résultat de l'identification est correct [44].

Dans le cas d'une vérification, la mesure de performance standard d'un système biométrique utilise les métriques énumérées ci-dessous.

Le Taux de Faux Rejets (TFR) ou False Reject Rate (FRR) : il exprime le pourcentage d'utilisateurs rejetés alors qu'ils devraient être acceptés par le système.

$$\text{TFR} = (\text{nombre des clients rejetés}) / (\text{nombre total d'accès clients}) \quad (1.1)$$

Le Taux de Fausses Acceptations (TFA) ou False Accept Rate (FAR) : il exprime le pourcentage d'utilisateurs acceptés par le système alors qu'ils devraient être rejetés.

$$\text{TFA} = (\text{nombre des imposteurs acceptés}) / (\text{nombre total d'accès imposteurs}) \quad (1.2)$$

Le Taux d'Égale Erreur (TEE) ou Equal Error Rate (EER) : Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où $\text{TFR} = \text{TFA}$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations [45].

La courbe ROC (Receiver Operating Characteristic) : est une démonstration visuelle du compromis entre les TFA et le TFR liées à un seuil variable.

1.6.2 Mesure des performances la reconnaissance faciale de la parenté

Les performances de vérification de la parenté sont calculées à partir de paramètre de précision [46].

$$\text{Précision} = 100 - \min(\text{TFP} + \text{TFN}) \quad (1.3)$$

TFP (Taux de Faux Positif) : ce taux représente le pourcentage des images déclarées parent par le système, mais elles sont en réalité des images non parent.

TFN (Taux de Faux Négatif) : ce taux représente le pourcentage des images déclarées non parent par le système, mais elles sont en réalité des images parent.

1.7 Articulation de la thèse

Le manuscrit de la thèse s'articule autour de six chapitres :

Le premier chapitre est une partie introductive au contexte de cette thèse consacré à la présentation des objectifs et des contributions de notre travail. Une description générale des différentes modalités de biométrie et la reconnaissance automatique de visage sont décrites. Dans lesquelles, nous avons présenté le détail de ces modalités biométriques ainsi que la place du visage parmi celles-ci. Ensuite, nous avons introduit la structure générale et le principe de fonctionnement d'un système de reconnaissance automatique de visage et de parenté, ses

différents défis scientifiques dans les environnements non contrôlés et les paramètres d'évaluation ainsi que les mesures des performances de ces systèmes.

Le deuxième chapitre est consacré à la présentation des fondamentaux d'apprentissage de sous-espaces multilinéaires. Nous présentons, dans un premier temps, le contexte et les outils essentiels de l'algèbre multilinéaire, Ensuite, nous discutons les trois types de base de projections multilinéaires : projection de vecteur en vecteur (projection linéaire), projection de tenseur en tenseur et la projection de tenseur en vecteur. En outre, leurs connexions sont analysées. Enfin, nous présentons le problème d'apprentissage de sous-espaces multilinéaires. En résumé, ce chapitre propose un traitement systématique des concepts fondamentaux nécessaires au développement de nouveaux algorithmes d'apprentissage de sous-espaces multilinéaires.

Le troisième chapitre est consacré à l'état de l'art de la reconnaissance de visage par l'analyse multidimensionnelle. Tout d'abord, nous présentons les techniques d'apprentissage de sous-espaces multilinéaires non supervisées. Ensuite, nous présentons les techniques d'apprentissage de sous-espaces multilinéaire supervisées. Le but de ce chapitre est d'évaluer les capacités et les lacunes d'algorithmes de reconnaissance de visage en cours afin d'identifier les problèmes et les orientations pour la recherche dans ce domaine.

Dans le quatrième chapitre nous proposons une nouvelle méthode de réduction de dimensionnalité multilinéaire supervisée MEFDA pour la reconnaissance faciale 2D et 3D qui agit directement sur les données tensorielles. Tout d'abord, la formulation mathématique de la méthode proposée est introduite. Ensuite, la méthodologie dans le cadre de la vérification de visage basé sur MEFDA est illustrée. Enfin, une étude approfondie visant à évaluer MEFDA par rapport à d'autres méthodes existantes, linéaires (avec 1-dimension) et multilinéaires (avec N-dimensions) est présentée et discutée. Trois ensembles d'expériences sont conçues et réalisées à l'aide des trois bases de données différentes, FRGC v2.0, CASIA 3D et Bosphorus.

Dans le cinquième chapitre nous introduisons une nouvelle méthode de réduction de dimensionnalité multilinéaire faiblement supervisée MSIDA pour la reconnaissance de l'identité et la parenté, robuste et efficace aux variations des milieux non contrôlés. Tout d'abord, la formulation mathématique de la méthode proposée est introduite. Ensuite, la méthodologie dans le cadre de la vérification de la parenté basée sur MSIDA est illustrée. Enfin, une validation expérimentale de la méthode proposée à l'aide de quatre bases de données

de visages différentes (LFW, Cornell KinFace, UB KinFace et TSkinFace) et la discussion des résultats sont données à la fin de ce chapitre.

Dans le sixième chapitre, nous concluons ce mémoire de thèse en résumant les points principaux de nos contributions et nous proposons des perspectives envisageables à explorer suite à notre travail.

Chapitre 2

Propriétés de l’algèbre multilinéaire

2.1 Introduction	19
2.2 Bases d'algèbre multilinéaire.....	19
2.2.1 Notations.....	19
2.2.2 Définition mathématique	21
2.2.2.1 Vectorisation : transformation tenseur en vecteur	21
2.2.2.2 Déploiement de tenseurs	22
2.2.2.3 Produit intérieur tensoriel et norme de Frobenius.....	23
2.2.2.4 Produit extérieur.....	23
2.2.2.5 Produit mode k : multiplication tenseur matrice.....	24
2.2.2.6 Produit de Kronecker	25
2.2.2.7 Mesure de distance tenseur	26
2.3 Projections multilinéaires	28
2.3.1 Projection vecteur en vecteur	28
2.3.2 Projection tenseur en tenseur	29
2.3.2 Projection tenseur en vecteur.....	31
2.4 Relations entre les projections multilinéaires.....	32
2.4.1 Nombre de paramètres à estimer	33
2.5 Apprentissage de sous-espace linéaire	34
2.5.1 Définition du problème.....	35
2.6 Apprentissage de sous-espace multilinéaire.....	36
2.6.1 Définition du problème.....	36
2.6.2 Mesures de dispersion pour les tenseurs.....	36
2.7 Réduction de la dimensionnalité par l’apprentissage de sous-espace	38
2.8 Conclusion.....	40

2.1 Introduction

Les données multidimensionnelles sont générées quotidiennement dans de nombreuses applications. Ceci conduit à une forte demande d'utilisation des outils d'algèbre multilinéaire. Le but de ce chapitre est de présenter les bases d'algèbre multilinéaire qui seront utiles dans l'élaboration des techniques présentées dans cette thèse et aider les lecteurs à comprendre les concepts multilinéaires. Dans la première partie de ce chapitre, nous présentons, au travers des propriétés d'algèbre multilinéaire, des outils nécessaires à la compréhension de l'approche tensorielle et nous définissons ce que nous appelons « tenseur » avec un tel formalisme. Ensuite, nous étudions les trois types de base de projections multilinéaires, et leurs connexions sous-jacentes. Enfin, nous présentons le problème d'apprentissage de sous-espace selon les deux catégories (linéaire/multilinéaire), et nous examinons la réduction de la dimensionnalité par les techniques de sous-espaces multilinéaires.

2.2 Bases d'algèbre multilinéaire

Cette section passe en revue les notations, les définitions et les concepts fondamentaux de l'algèbre multilinéaire, qui sont nécessaires à la définition du problème d'apprentissage de sous-espaces multilinéaires pour les données multidimensionnelles.

2.2.1 Notations

Cette thèse suit les conventions de notation dans la littérature sur l'algèbre multilinéaire, la reconnaissance des formes et l'apprentissage adaptatif [47-49]. Les scalaires sont désignés par des lettres minuscules et majuscules (i, j, L, N, \dots), les vecteurs par des lettres minuscules en gras ($\mathbf{x}, \mathbf{y}, \dots$), les matrices par des lettres majuscules italiques (V, Y, \dots), et les tenseurs par des lettres majuscules en gras ($\mathbf{A}, \mathbf{B}, \dots$).

L'algèbre multilinéaire est la base de l'analyse des données tensorielles étudiée en mathématiques depuis plusieurs décennies [47, 48] [50]. Un tenseur est considéré comme un tableau multidimensionnel [51-53]. C'est l'extension d'ordre élevé du vecteur et de la matrice à N entrées. Un tenseur d'ordre N est représenté par: $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, donc il décrit par N indices, où $I_K, 1 \leq K \leq N$, indique la dimension du mode k du tenseur. Chaque élément du tenseur \mathbf{A} est noté par $a_{i_1 i_2 \dots i_N}$. Les figures 2.1 (a), 2.1 (b) et 2.1 (c), présentent des exemples d'un tenseur d'ordre 2, 3 et 4, respectivement.

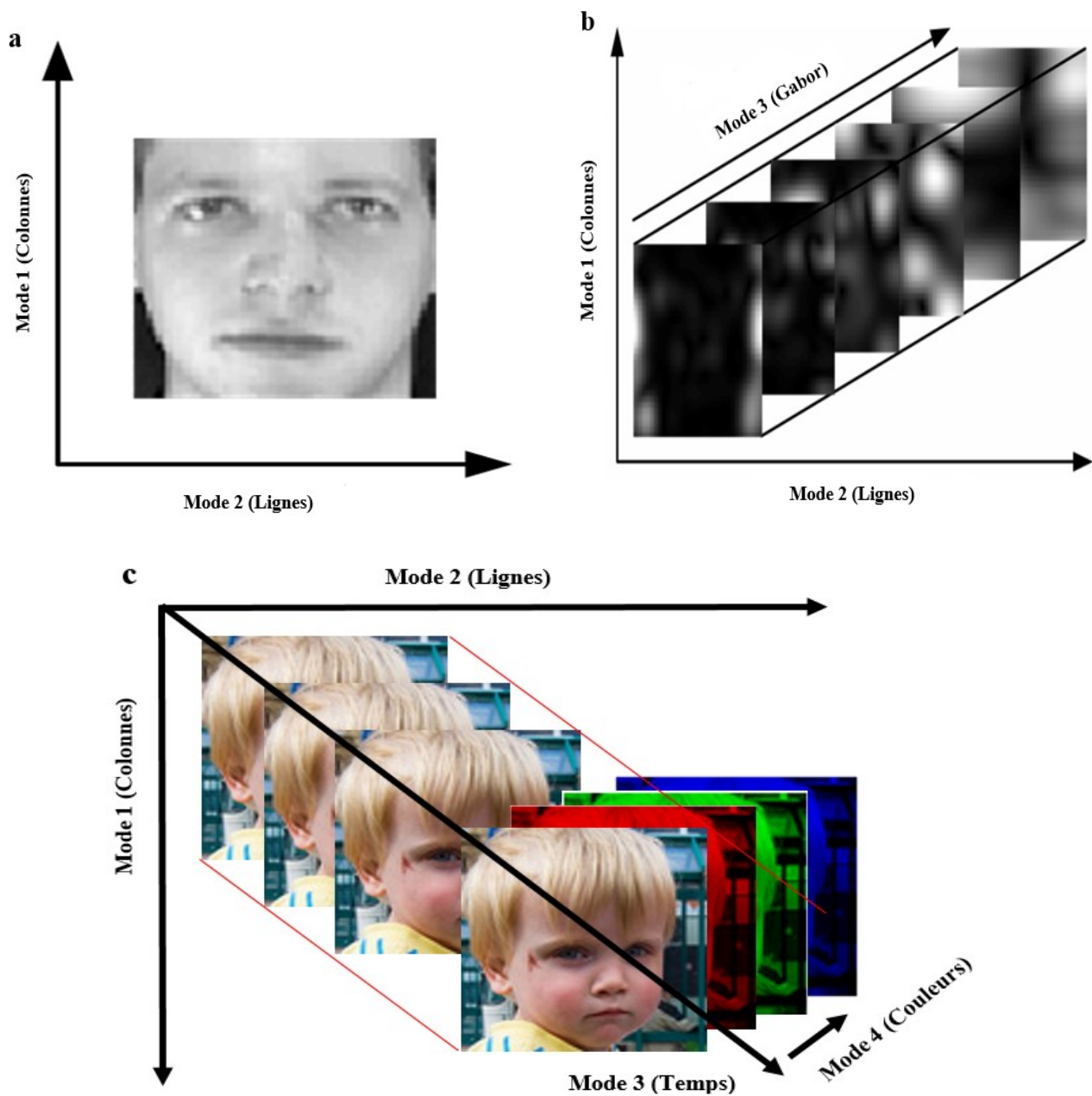


Figure 2.1: Exemple d'une représentation tensorielle : (a) deuxième ordre, (b) troisième ordre, (c) quatrième ordre (Séquence vidéo couleur).

L'espace du tenseur d'ordre N est composé de N mode de sous-espaces. Un ensemble de données comprenant N paramètres peut-être modéliser comme un tableau de N entrées, dans lequel chaque mode du tenseur est associé à un paramètre différent. Le tableau 2.1 donne quelques exemples de la modélisation tensorielle.

Tableau 2.1: Exemples de modélisations tensorielles.

Modélisations tensorielles	Exemple
Tenseur d'ordre 0	Longueur d'onde
Tenseur d'ordre 1	Coordonnées cartésiennes
Tenseur d'ordre 2	Image en niveau de gris
Tenseur d'ordre 3	Image RGB (couleur)
Tenseur d'ordre 4	Séquence vidéo couleur

2.2.2 Définition mathématique

Nous rappelons ici quelques définitions essentielles de l'algèbre multilinéaire qui seront utiles dans la suite de cette thèse.

2.2.2.1 Vectorisation : transformation tenseur en vecteur

Similaire à la vectorisation d'une matrice, la vectorisation d'un tenseur est une transformation linéaire qui convertit le tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, en un vecteur colonne $\mathbf{a} \in \mathbb{R}^{\prod_{n=1}^N I_n}$, dénoté par $\mathbf{a} = \text{vec}(\mathbf{A})$. Les figures 2.2 (b), 2.2 (c), et 2.2 (d) donnent des illustrations visuelles des vecteurs du tenseur de troisième ordre \mathbf{A} de la figure 2.2 (a) selon les modes 1, 2 et 3.

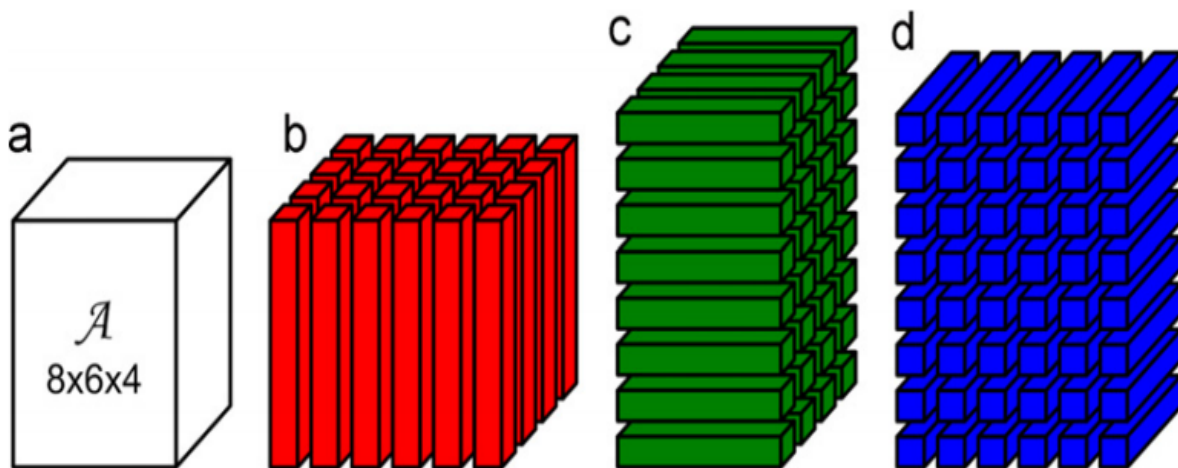


Figure 2.2: Illustration visuelle des vecteurs en mode k : (a) tenseur $\mathbf{A} \in \mathbb{R}^{8 \times 6 \times 4}$, (b) vecteurs mode 1, (c) vecteurs mode 2, (d) vecteurs mode 3 [54].

2.2.2.2 Déploiement de tenseurs

Le traitement des données multidimensionnelles peut être dans un mode privilégié. À cette fin, nous utilisons les matrices de déploiement dans le mode k pour travailler avec les tenseurs selon un mode favorisé. De ce fait, le déploiement dans le mode k du tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ dans une matrice $A^{(k)} \in \mathbb{R}^{I_k \times \prod_{i=1, i \neq k}^N I_i}$ est noté par :

$$A^{(k)} \leftarrow_k \mathbf{A} \tag{2.1}$$

Où les entrées sont données par :

$$A_{i_k j}^{(k)} = \mathbf{A}_{i_1 \dots i_N}, \quad j = 1 + \sum_{q=1, q \neq k}^N (i_q - 1) \prod_{p=1+1, p \neq k}^N I_p \tag{2.2}$$

Le mode k matrices est formé en arrangeant tous les modes k vecteurs séquentiellement ainsi que ses colonnes. La figure 2.3 montre le déploiement du tenseur \mathbf{A} de la Figure. 2.2 (a) dans les différents modes.

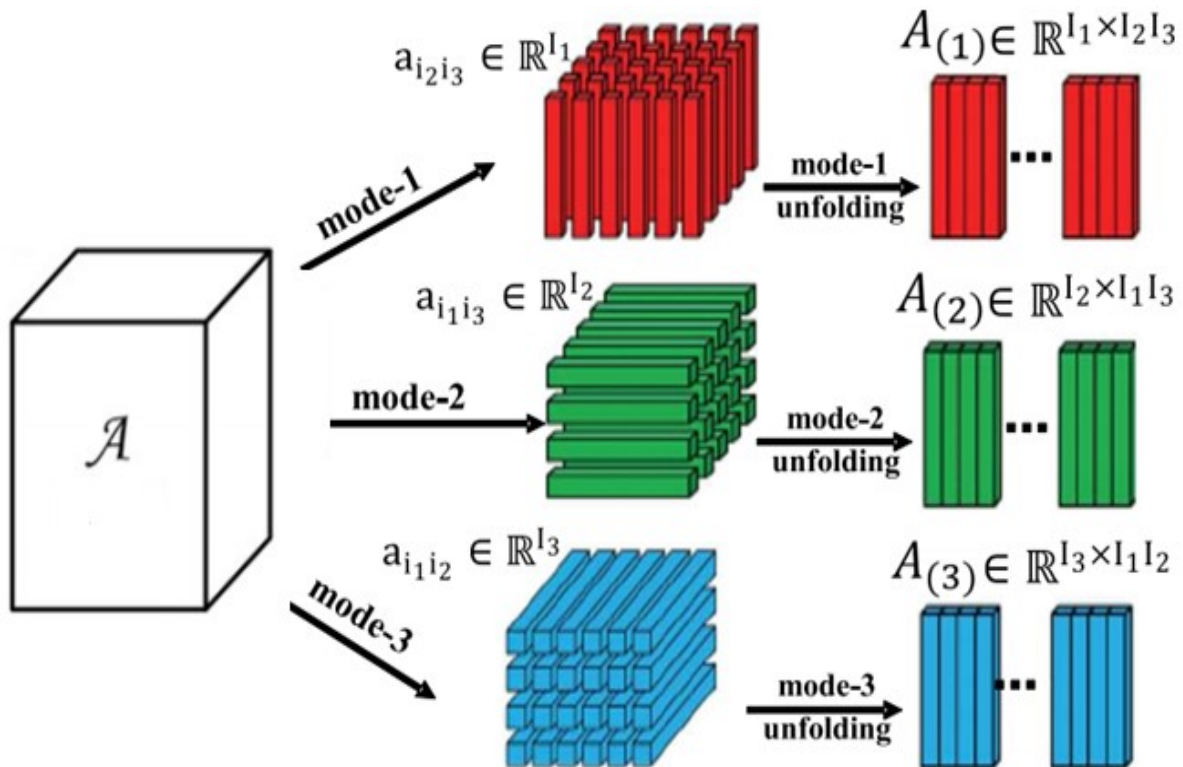


Figure 2.3: Déploiement d'un tenseur d'ordre 3 dans les différents modes [54].

2.2.2.3 Produit intérieur tensoriel et norme de Frobenius

Considérons deux tenseurs avec le même ordre et de même dimension $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. Le produit intérieur entre ces deux tenseurs est donné par :

$$\langle \mathbf{A} | \mathbf{B} \rangle = \sum_{i_1, \dots, i_N} a_{i_1 \dots i_N} \cdot b_{i_1 \dots i_N} \quad (2.3)$$

Propriété 2.1:

$$\langle \mathbf{A} | \mathbf{B} \rangle = \text{tr} (A_k B_k^T) \quad (2.4)$$

Où tr est l'opération trace d'une matrice.

En outre, la **norme de Frobenius** d'un tenseur \mathbf{A} est défini comme :

$$\|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \dots i_N}^2} \quad (2.5)$$

Où :

$$\|\mathbf{A}\|^2 = \sum_{i_1, \dots, i_N} a_{i_1 \dots i_N}^2 = \text{tr} (A_n A_n^T) \quad (2.6)$$

2.2.2.4 Produit extérieur

Le produit extérieur de plusieurs vecteurs $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(N)}$ de dimension I_1, \dots, I_N et appartenant respectivement aux espaces vectoriels $E^{(1)}, \dots, E^{(N)}$, détermine le tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ de rang un où :

$$\mathbf{A} = \mathbf{v}^{(1)} \circ \mathbf{v}^{(2)} \circ \dots \circ \mathbf{v}^{(N)} \quad (2.7)$$

Ce qui signifie que

$$\mathbf{A}(i_1, i_2, \dots, i_N) = \mathbf{v}^{(1)}(i_1) \cdot \mathbf{v}^{(2)}(i_2) \cdot \dots \cdot \mathbf{v}^{(N)}(i_N) \quad (2.8)$$

L'élément $a_{i_1 \dots i_N}$ de tenseur \mathbf{A} est défini par le produit :

$$a_{i_1 \dots i_N} = \prod_{n=1}^N v_{i_n}^{(N)} \quad (2.9)$$

Où $v_{i_n}^{(N)}$ est la $i_n^{\text{ième}}$ composante du vecteur $\mathbf{v}^{(N)}$

2.2.2.5 Produit mode k : multiplication tenseur matrice

Ce produit peut s'interpréter comme un produit scalaire sur le mode k. Le produit mode k d'un tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ par une matrice $G \in \mathbb{R}^{J_k \times I_k}$ ($K=1, \dots, N$) est noté $\mathbf{A} \times_k G$ et donne un tenseur \mathbf{B} tel que :

$$\mathbf{B} = \mathbf{A} \times_k G \in \mathbb{R}^{I_1 \times \dots \times I_{k-1} \times J_k \times J_{k+1} \times \dots \times I_N} \quad (2.10)$$

Où chaque élément du tenseur est donné par :

$$\mathbf{B} = (\mathbf{A} \times_k G)_{i_1 i_2 \dots i_{k-1} j_k j_{k+1} \dots i_N} = \sum_{i_k=1}^{I_k} a_{i_1 i_2 \dots i_{k-1} i_k j_{k+1} \dots i_N} \cdot g_{j_k i_k} \quad (2.11)$$

Ce résultat, peut s'écrire avec la forme matricielle en utilisant les matrices dépliantes :

$$B_k = G \cdot A_k \quad (2.12)$$

Où A_k et B_k représentent les matrices dépliantes des tenseurs \mathbf{A} et \mathbf{B} dans le mode k. Ainsi le tenseur résultant des N mode k produits entre le tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ et les matrices $G^{(k)} \in \mathbb{R}^{J_k \times I_k}$, $k = 1, \dots, N$ se définit par :

$$\mathbf{B} = \mathbf{A} \times_1 G^{(1)} \dots \times_N G^{(N)} \in \mathbb{R}^{J_1 \times \dots \times J_N} \quad (2.13)$$

Les éléments indice (J_1, \dots, J_N) du tenseur \mathbf{B} sont donnés par :

$$b_{j_1 \dots j_N} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} a_{i_1 \dots i_N} g_{j_1 i_1}^{(1)} \dots g_{j_N i_N}^{(N)} \quad (2.14)$$

Où $i_k \in \{1, \dots, I_k\}$ et $J_k \in N^*$, $k = 1, \dots, N$

La figure 2.4 illustre un exemple de produit mode 1 du tenseur de troisième ordre $\mathbf{X} \in \mathbb{R}^{12 \times 7 \times 4}$ avec matrice $G^{(T)} \in \mathbb{R}^{4 \times 12}$, où le résultat est un tenseur $\mathbf{X} \in \mathbb{R}^{4 \times 7 \times 4}$

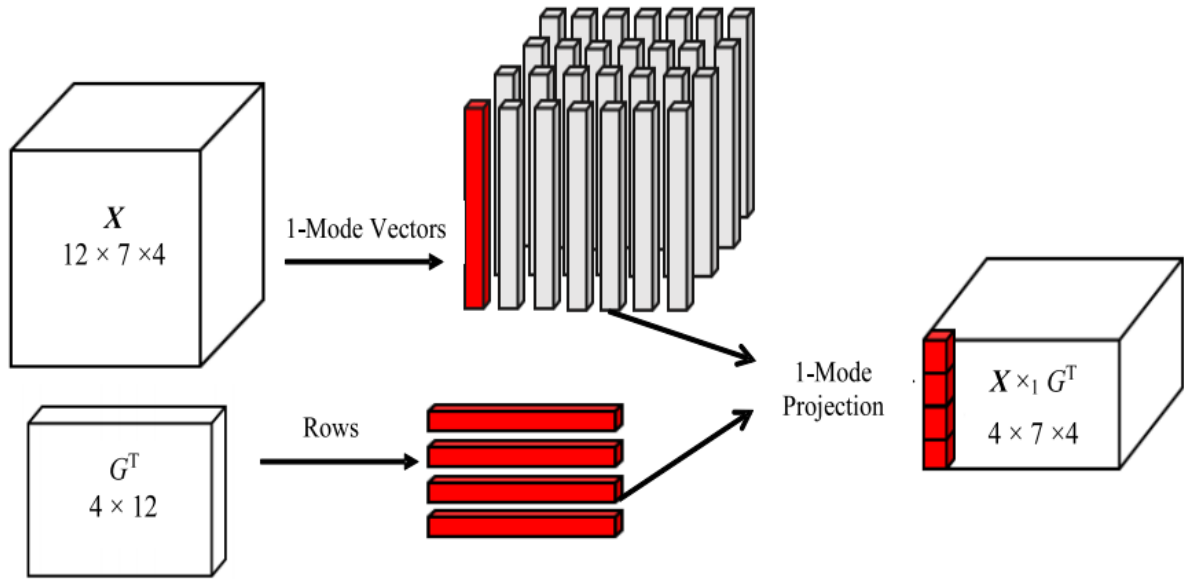


Figure 2.4: Illustration visuelle d'un produit mode 1 du tenseur X du troisième ordre avec une matrice G .

Propriété 2.2 :

Donné un tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ et deux matrices $G \in \mathbb{R}^{J_k \times I_k}$ et $V \in \mathbb{R}^{J_n \times I_n}$, où $n \neq k$, nous avons

$$(\mathbf{A} \times_n G) \times_k V = (\mathbf{A} \times_k V) \times_n G \quad (2.15)$$

Propriété 2.3 :

Donné un tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ et deux matrices $G \in \mathbb{R}^{I_k \times I_k}$ et $V \in \mathbb{R}^{J_k \times I_k}$, nous avons

$$(\mathbf{A} \times_k G) \times_k V = \mathbf{A} \times_k (V \cdot G) \quad (2.16)$$

2.2.2.6 Produit de Kronecker

Le produit de Kronecker [55] de deux matrices $G \in \mathbb{R}^{I_1 \times I_2}$ et $V \in \mathbb{R}^{J_1 \times J_2}$ est noté $G \otimes V$ et conduit à une matrice de dimension $(I_1 J_1 \times I_2 J_2)$ définie par :

$$G \otimes V = \begin{pmatrix} g_{11}V & \dots & g_{1I_2}V \\ \vdots & \vdots & \vdots \\ g_{I_1 1}V & \dots & g_{I_1 I_2}V \end{pmatrix} \quad (2.17)$$

Ce produit n'est pas commutatif, cependant il est associatif et distributif par rapport à l'addition. Voici les deux propriétés relatives au produit de Kronecker :

Propriété 2.4 :

Considérons les matrices $G_k \in \mathbb{R}^{I_k \times I_k}$, $k \in \{1, \dots, N\}$, et $V_k \in \mathbb{R}^{J_k \times P_k}$, $k \in \{1, \dots, N\}$. Quel que soit $N \in \mathbb{N}^*$, le produit de Kronecker \otimes vérifie la propriété suivante :

$$(G_1 \otimes G_2 \dots \otimes G_N)(V_1 \otimes V_2 \dots \otimes V_N) = (G_1 V_1) \otimes (G_2 V_2) \dots \otimes (G_N V_N) \quad (2.18)$$

Propriété 2.5 :

La transposée d'un produit de Kronecker de plusieurs matrices, quelles que soient les matrices $G_k \in \mathbb{R}^{J_k \times I_k}$ ($k \in \{1, \dots, N\}$), vérifie la relation suivante :

$$(G_1 \otimes G_2 \dots \otimes G_N)^T = (G_1^T \otimes G_2^T \dots \otimes G_N^T) \quad (2.19)$$

2.2.2.7 Mesure de distance tenseur

La norme de Frobenius peut être utilisée pour mesurer la distance entre les tenseurs **A** et **B** comme

$$\text{dist}(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|_F \quad (2.20)$$

Bien qu'il s'agisse d'une mesure basée sur le tenseur, il peut être prouvé qu'elle est équivalente à une mesure de distance des représentations vectorielles correspondantes, comme démontré dans [56]. Soit $\text{vec}(\mathbf{A})$ la représentation vectorielle (vectorisation) de **A**, deux propriétés concernant le produit intérieur entre deux tenseurs est obtenue comme suit :

Propriété 2.6 :

$$\langle \mathbf{A} | \mathbf{B} \rangle = \langle \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B}) \rangle = [\text{vec}(\mathbf{B})]^T \text{vec}(\mathbf{A}) \quad (2.21)$$

À partir de l'équation (2.3),

$$\begin{aligned} \langle \mathbf{A} | \mathbf{B} \rangle &= \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} \mathbf{A}(i_1, i_2, \dots, i_N) \cdot \mathbf{B}(i_1, i_2, \dots, i_N) \\ &= \sum_{i=1}^{\prod_{n=1}^N I_n} \text{vec}(\mathbf{A})(i) \cdot \text{vec}(\mathbf{B})(i) \end{aligned}$$

$$\begin{aligned}
&= \langle \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B}) \rangle \\
&= [\text{vec}(\mathbf{B})]^T \text{vec}(\mathbf{A})
\end{aligned}$$

Ensuite, il est simple de montrer que

Propriété 2.7 :

$$\text{dist}(\mathbf{A}, \mathbf{B}) = \|\text{vec}(\mathbf{A}) - \text{vec}(\mathbf{B})\|_2 \quad (2.22)$$

À partir de la propriété (2.6),

$$\begin{aligned}
\text{dist}(\mathbf{A}, \mathbf{B}) &= \|\mathbf{A} - \mathbf{B}\|_F \\
&= \sqrt{\langle \mathbf{A} - \mathbf{B}, \mathbf{A} - \mathbf{B} \rangle} \\
&= \sqrt{\langle \text{vec}(\mathbf{A}) - \text{vec}(\mathbf{B}), \text{vec}(\mathbf{A}) - \text{vec}(\mathbf{B}) \rangle} \\
&= \|\text{vec}(\mathbf{A}) - \text{vec}(\mathbf{B})\|_2
\end{aligned}$$

La propriété 2.7 indique que la norme de Frobenius de la différence entre deux tenseurs est égale à la distance euclidienne entre leurs représentations vectorisées. La norme de tenseur de Frobenius est une mesure ponctuelle [57] sans prendre en compte la structure du tenseur.

Les classificateurs utilisent fréquemment les mesures de distance pour mesurer la similarité ou la dissimilarité. En outre, il est possible de concevoir des classificateurs en tenant compte de la représentation ou de la structure matrice / tenseur. Par exemple, Wang et al [58] ont proposé un classifieur spécialement conçu pour les représentations matricielles des motifs et ont montré qu'un tel classifieur présente des avantages pour les caractéristiques extraites des représentations matricielles.

2.3 Projections multilinéaires

Un sous-espace multilinéaire est défini par une projection multilinéaire qui mappe les données d'entrée d'un espace à un autre espace (dimension inférieure) [59]. Par conséquent, la projection multilinéaire est un concept important à saisir avant de passer à l'apprentissage multilinéaire de sous-espace.

Cette section présente les trois projections multilinéaires de base nommées en fonction des représentations en entrée et en sortie de la projection: la projection traditionnelle vecteur en vecteur, la projection tenseur en tenseur et la projection tenseur en vecteur. De plus, nous étudions les relations entre ces projections.

2.3.1 Projection vecteur en vecteur

La projection linéaire est une transformation standard largement utilisée dans diverses applications [60, 61]. Une projection linéaire prend un vecteur \mathbf{x} et le projette sur un autre vecteur \mathbf{y} en utilisant une matrice de projection $U \in \mathbb{R}^{I \times P}$

$$\mathbf{y} = U^T \mathbf{x} = \mathbf{x} \times_1 U^T \quad (2.23)$$

Dans les applications typiques de reconnaissance de formes, $P \ll I$. Par conséquent, la projection linéaire est une projection de vecteur en vecteur car il mappe un vecteur sur un autre vecteur. Lorsque les données d'entrée sont une matrice ou un tenseur d'ordre supérieur \mathbf{A} , il doit être vectorisé comme $\mathbf{x} = \text{vec}(\mathbf{A})$ (transformé en un vecteur) avant la projection. La figure 2.5 illustre la projection de vecteur en vecteur d'un objet tenseur \mathbf{A} .

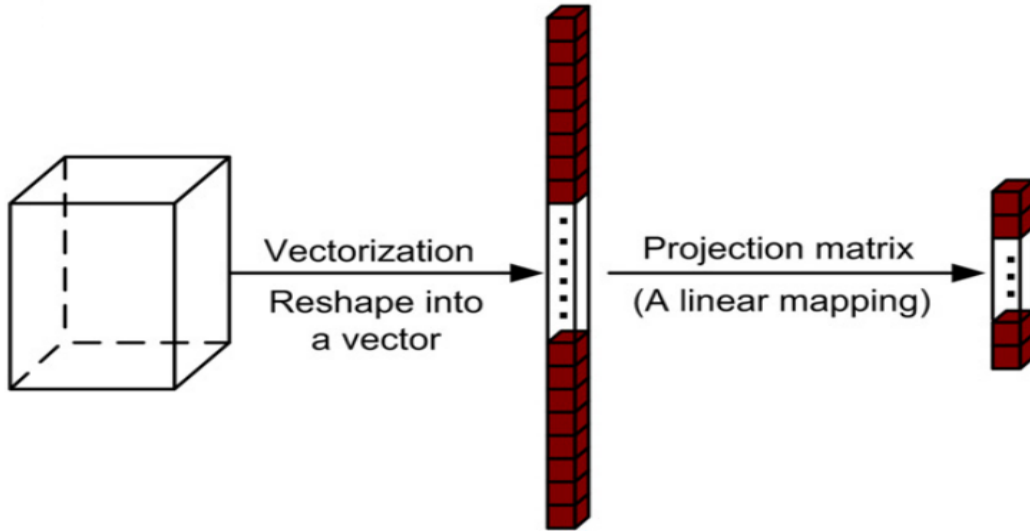


Figure 2.5: Illustration de la projection de vecteur en vecteur [54].

Outre la projection linéaire traditionnelle, il existe d'autres moyens de projeter un tenseur dans un espace de faible dimension, comme illustré à la figure 2.6, qui sera présenté ci-dessous.

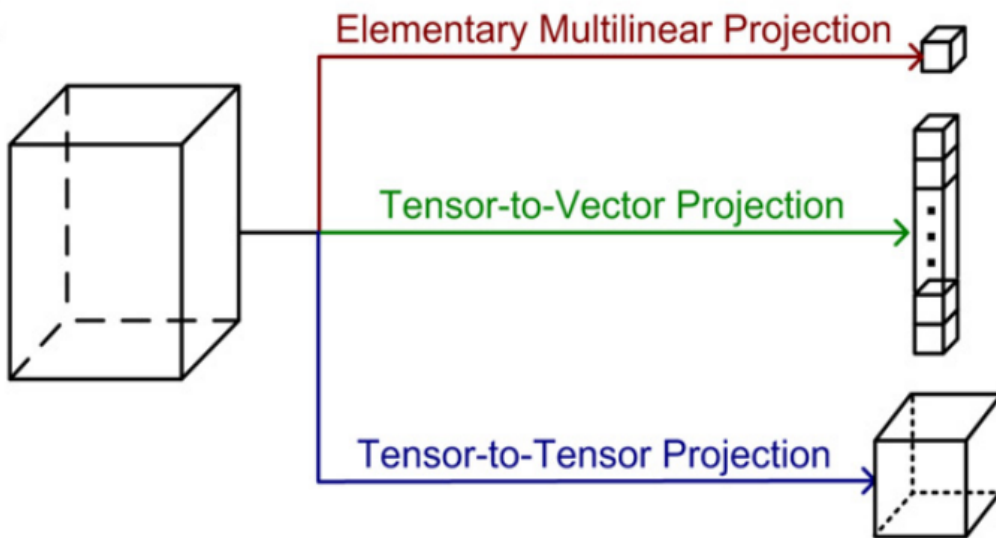


Figure 2.6: Façon alternative de projeter un tenseur dans un espace de petite dimension [54].

2.3.2 Projection tenseur en tenseur

En plus de la projection traditionnelle de vecteur en vecteur, un tenseur peut également être projeté sur un autre tenseur (généralement du même ordre), appelé projection de tenseur en

tenseur. Un tenseur \mathbf{X} d'ordre N situé dans l'espace tensoriel $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$ [47, 62]. Ainsi, l'espace tensoriel peut être considéré comme le produit de Kronecker de N espaces vectoriels (linéaires) $\mathbb{R}^{I_1}, \mathbb{R}^{I_2}, \dots, \mathbb{R}^{I_N}$ [62]. Pour la projection d'un tenseur \mathbf{X} dans un espace tensoriel $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$ à un autre tenseur \mathbf{Y} dans un espace tensoriel de dimension inférieure $\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \dots \otimes \mathbb{R}^{P_N}$, où $P_k \leq I_k$ pour tout k , N matrices de projection sont utilisés [48]

$$\mathbf{Y} = \mathbf{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \dots \times_N U^{(N)T} \tag{2.24}$$

Cela peut être fait en N étapes, où à la $n^{\text{ième}}$ étape, chaque vecteur en mode k est projeté dans une dimension inférieure P_k par $U^{(k)}$, comme le présente la figure 2.7 (a), Cette projection multilinéaire peut être réalisée à travers N multiplications en mode k . La figure 2.7 (b) montre la projection d'un tenseur en mode 1 en utilisant une matrice de projection en mode 1, qui projette chaque vecteur en mode 1 du tenseur d'origine en un vecteur de faible dimension.

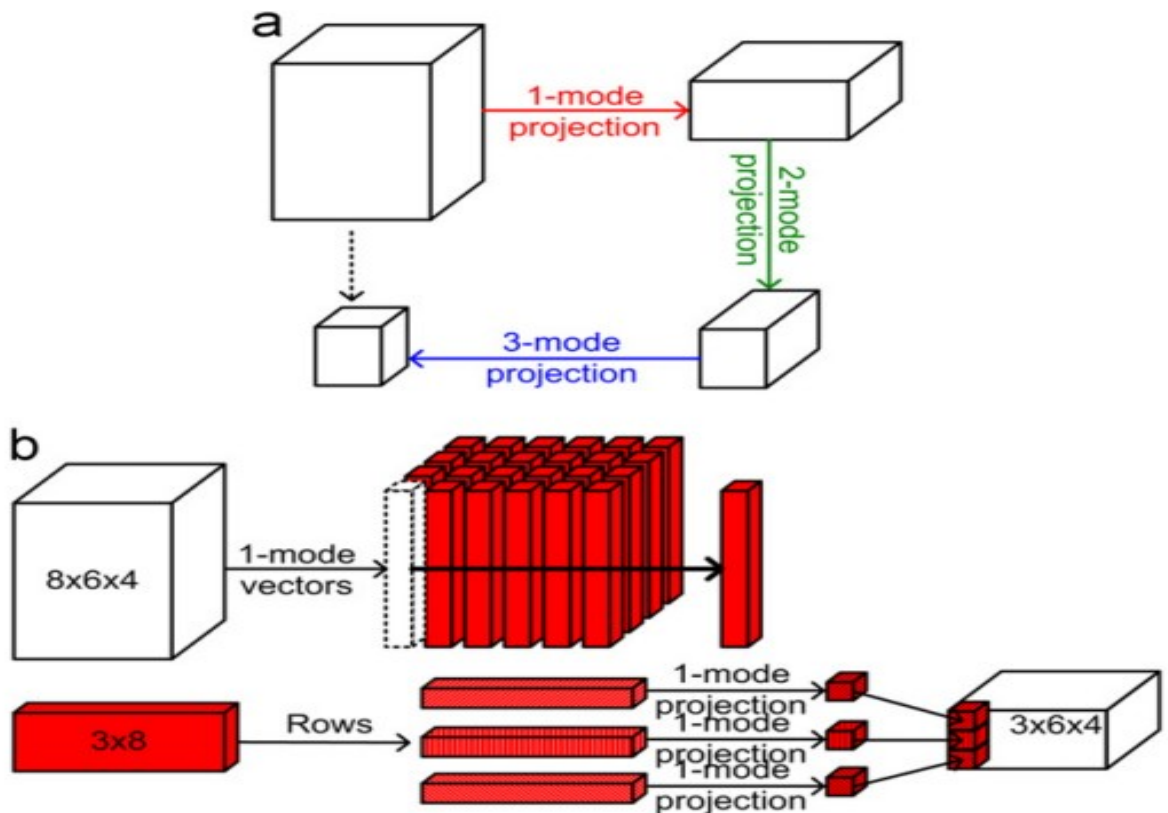


Figure 2.7: Illustration de la projection de tenseur en tenseur: (a) projection d'un tenseur dans tous les modes (b) projection d'un tenseur dans un seul mode [54].

2.3.2 Projection tenseur en vecteur

Ensuite, une troisième projection multilinéaire est introduite, d'un espace tenseur à un espace vectoriel, appelée projection de tenseur en vecteur, que l'on appelle les projections de rang un dans certains travaux [63-65]. La projection de tenseur en vecteur projette un tenseur sur un vecteur et peut être visualisée sous la forme de projections multiples d'un tenseur en un scalaire, comme illustré à la figure 2.8 (a), où la projection de tenseur en vecteur d'un tenseur $\mathbf{A} \in \mathbb{R}^{8 \times 6 \times 4}$ à un vecteur $\mathbf{p} \times 1$ consiste en P projections de \mathbf{A} à un scalaire. Ainsi, la projection d'un tenseur sur un scalaire est considérée en premier.

Un tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ peut être projeté sur un point y à travers N unités de vecteurs de projection $\{\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T}, \dots, \mathbf{u}^{(N)T}\}$ comme

$$y = \mathbf{A} \times_1 \mathbf{u}^{(1)T} \times_1 \mathbf{u}^{(2)T} \dots \times_N \mathbf{u}^{(N)T}, \quad \|\mathbf{u}^{(n)}\| = 1 \quad \text{Pour } n = 1, \dots, N \quad (2.25)$$

Où $\|\cdot\|$ est la norme euclidienne pour les vecteurs. Il peut être écrit dans le produit scalaire (2.3) comme suit:

$$y = \langle \mathbf{A}, \mathbf{u}^{(1)} \circ \mathbf{u}^{(1)} \circ \dots \circ \mathbf{u}^{(N)} \rangle \quad (2.26)$$

En plaçant $\langle \mathbf{u}^{(1)} \circ \mathbf{u}^{(1)} \circ \dots \circ \mathbf{u}^{(N)} \rangle$, alors $y = \langle \mathbf{A}, \mathbf{U} \rangle$. Cette projection multilinéaire $\{\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T}, \dots, \mathbf{u}^{(N)T}\}$ est nommée projection élémentaire multilinéaire (voir figure 2.6), la projection d'un tenseur sur une seule ligne (résultant d'un scalaire) et consiste en un vecteur de projection dans chaque mode. La figure 2.8 (b) illustre une projection élémentaire multilinéaire d'un tenseur $\mathbf{A} \in \mathbb{R}^{8 \times 6 \times 4}$. La projection tenseur en vecteur d'un objet tenseur \mathbf{X} sur un vecteur \mathbf{y} dans une l'espace vectoriel de dimension P est constitué de P projections multilinéaires élémentaires.

$$\{\mathbf{u}_p^{(1)T}, \mathbf{u}_p^{(1)T}, \dots, \mathbf{u}_p^{(N)T}\}, p=1, \dots, P \quad (2.27)$$

qui peut être écrit de manière concise $\{\mathbf{u}_p^{(n)T}, n = 1, \dots, N\}_{p=1}^P$. La projection de tenseur en vecteur de \mathbf{A} à \mathbf{y} s'écrit alors:

$$\mathbf{y} = \mathbf{X} \times_{n=1}^N \{\mathbf{u}_p^{(n)T}, n = 1, \dots, N\}_{p=1}^P \quad (2.28)$$

Où la $p^{\text{ième}}$ composante de y est obtenue à partir de la $p^{\text{ième}}$ projection multilinéaire élémentaire sous la forme

$$y(p) = \mathbf{X} \times_1 u_p^{(1)T} \times_2 u_p^{(2)T} \dots \times_1 u_p^{(N)T} \quad (2.29)$$

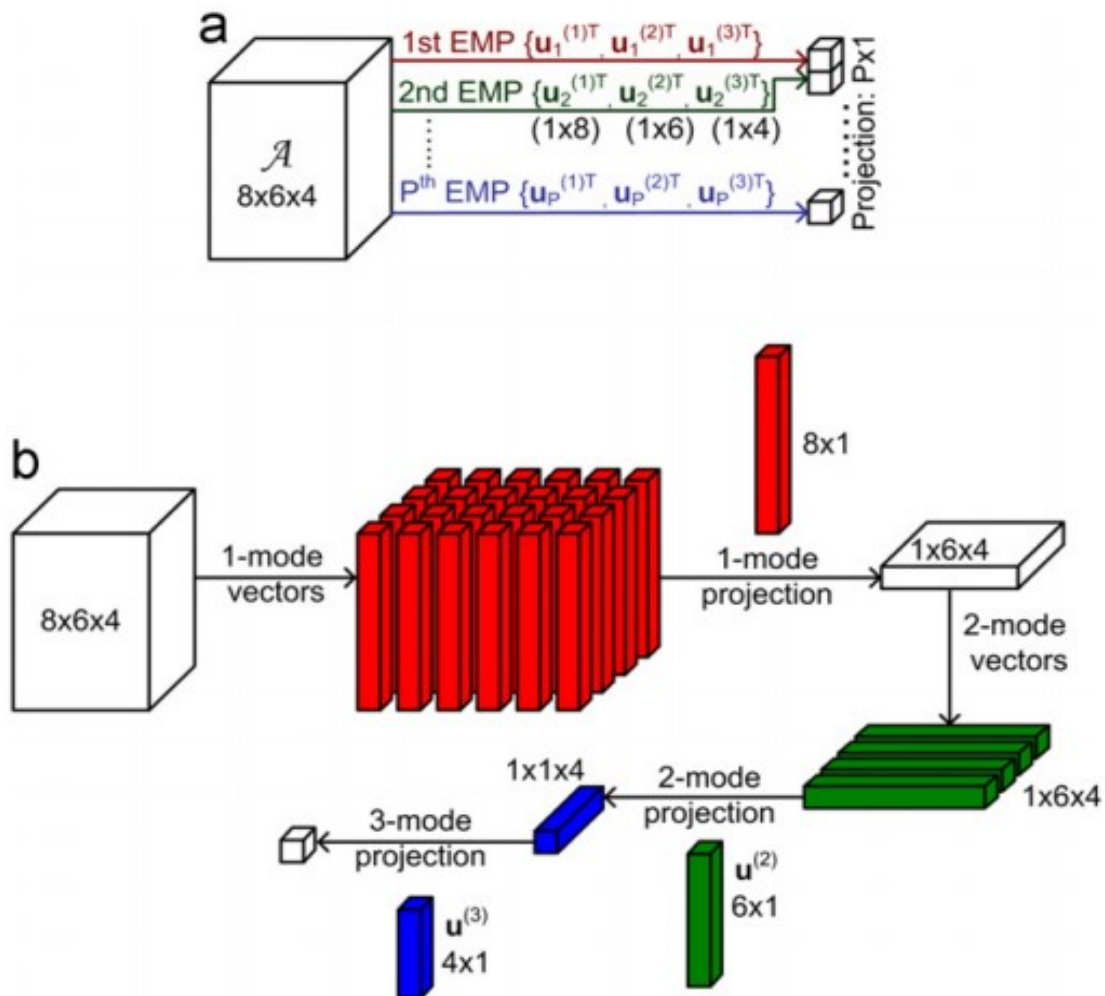


Figure 2.8: Illustration de la projection de tenseur en vecteur par projection élémentaire multilinéaire: (a) projection de tenseur en vecteur et (b) projection élémentaire multilinéaire [54].

2.4 Relations entre les projections multilinéaires

Avec l'introduction des trois projections multilinéaires de base, il est intéressant d'examiner leurs relations. Il est facile de vérifier que la projection de vecteur en vecteur est un cas particulier de la projection de tenseur en tenseur et de la projection de tenseur en vecteur avec

$N = 1$. La projection multilinéaire élémentaire est la version dégénérée de la projection de tenseur en tenseur avec $P_k = 1$ pour tout k . D'autre part, chaque élément projeté dans la projection de tenseur en tenseur peut être considéré comme la projection d'une projection élémentaire multilinéaire formée en prenant une colonne de chacune des matrices de projection. Ainsi, le tenseur projeté dans la projection du tenseur en tenseur est obtenu par $\prod_{k=1}^N P_k$ projections multilinéaires élémentaires interdépendantes, alors que dans la projection du tenseur en vecteur, les P projections multilinéaires élémentaires obtenues séquentiellement ne sont pas généralement interdépendantes.

De plus, la projection élémentaire multilinéaire $\{\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T}, \dots, \mathbf{u}^{(N)T}\}$ peut être écrite comme:

$$y = \langle \mathbf{A}, \mathbf{B} \rangle = \langle \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B}) \rangle = [\text{vec}(\mathbf{B})]^T \text{vec}(\mathbf{A}) \quad (2.29)$$

Par la propriété 2.6. Ainsi, une projection élémentaire multilinéaire est équivalente à une projection linéaire de $\text{vec}(\mathbf{A})$. Puisque $\mathbf{B} = \mathbf{b}^{(1)} \circ \mathbf{b}^{(2)} \circ \dots \circ \mathbf{b}^{(N)}$, l'équation (2.28) indique que la projection multilinéaire élémentaire est équivalente à une projection linéaire pour $P = 1$ avec une contrainte sur le vecteur de projection telle qu'il s'agisse de la représentation vectorisée d'un tenseur de rang un.

2.4.1 Nombre de paramètres à estimer

Le nombre de paramètres à estimer dans une projection particulière indique la complexité du modèle, une préoccupation importante dans la pratique. Comparé à un vecteur de projection de taille $I \times 1$ dans la projection du vecteur en vecteur spécifiée par I paramètres ($I = \prod_{n=1}^N I_n$) pour un tenseur d'ordre n , une projection élémentaire multilinéaire dans la projection du tenseur en vecteur peut être spécifiée par $\sum_{n=1}^N I_n$ paramètres. Par conséquent, projeter un tenseur de taille $\prod_{n=1}^N I_n$ en vecteur de taille $P \times 1$, la projection du tenseur en vecteur besoin d'estimer seulement $P \cdot \sum_{n=1}^N I_n$ paramètres, tandis que la projection du vecteur en vecteur besoin d'estimer $P \cdot \prod_{n=1}^N I_n$ paramètres. L'implication dans le problème de la reconnaissance des formes est que la projection tenseur en vecteur a moins de paramètres à estimer tout en étant plus contrainte pour les solutions, et la projection vecteur en vecteur a moins de contraintes sur les solutions recherchées tout en ayant plus de paramètres à estimer. En d'autres termes, la projection du tenseur en vecteur a un modèle plus simple que la projection du vecteur en vecteur. Pour la projection du tenseur en tenseur avec la même quantité de réduction de dimensionnalité,

$\sum_{n=1}^N P_n \times I_n$ paramètres doivent être estimés. Ainsi, en raison des vecteurs de projection partagés, la projection du tenseur en tenseur peut avoir besoin d'estimer encore moins de paramètres et son modèle peut être encore plus simple. Le tableau 2.2 compare le nombre de paramètres à estimer selon les trois projections pour une même réduction de dimensionnalité dans plusieurs cas. Le tableau ci-dessous montre que pour les tenseurs d’ordre supérieur, le modèle de projection classique du vecteur en vecteur devient extrêmement complexe et que l’estimation de paramètres devient extrêmement difficile. Cela conduit souvent au problème de la petite taille de l'échantillon dans la pratique lorsque le nombre d'échantillons d'apprentissage disponibles est limité.

Tableau 2.2: Nombre de paramètres à estimer selon trois projections multilinéaires [66].

Entrée	Sortie	Projection du vecteur en vecteur	Projection du tenseur en vecteur	Projection du tenseur en tenseur
$\prod_{n=1}^N I_n$	P	$P \cdot \prod_{n=1}^N I_n$	$P \cdot \sum_{n=1}^N I_n$	$\sum_{n=1}^N P_n \times I_n$
10 × 10	4	400	80	40 ($P_n = 2$)
100 × 100	4	40,000	800	400 ($P_n = 2$)
100 × 100 × 100	8	8,000,000	2400	600 ($P_n = 2$)
$\prod_{n=1}^4 100$	16	1,600,000,000	6400	800 ($P_n = 2$)

2.5 Apprentissage de sous-espace linéaire

Pour des raisons de simplicité de calcul et de concept, une représentation qui est une transformation linéaire des données d'origine est souvent préférée. Par conséquent, les algorithmes de dimensionnalité utilisant des méthodes d'apprentissage par sous-espace linéaire ont suscité un intérêt considérable ces dernières années [67-69]. Ces algorithmes résolvent une projection linéaire avec certains critères d'optimalité, à partir d'un ensemble d'échantillons d'apprentissage. Dans l’approche bien connue d’apprentissage par sous-espace linéaire, chaque image de l’espace de données est représentée sous forme de vecteur dans un espace de grande dimension. Cependant, toute image faciale est essentiellement un tenseur du deuxième ou du troisième ordre et doit donc être transformée en un vecteur en représentation vectorielle. Une technique typique utilisée pour mener à bien ce processus est ce que l’on appelle la vectorisation

matrice-vecteur qui concatène toutes les lignes ou colonnes de la matrice pour donner un long vecteur.

2.5.1 Définition du problème

Un ensemble de N échantillons vectoriels $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ est disponible pour l'apprentissage, où chaque échantillon \mathbf{v}_m est un vecteur $I \times 1$ dans un espace vectoriel \mathbb{R}^I . L'objectif d'apprentissage linéaire de sous-espace est de trouver une transformation linéaire (projection) $Y \in \mathbb{R}^{I \times P}$ tel que les échantillons du projet $\{\mathbf{w}_m = Y^T \mathbf{v}_m\}$ satisfaisent à un critère d'optimalité, où $\mathbf{w}_m \in \mathbb{R}^{P \times 1}$ et $P < I$ [70].

PCA [19] et LDA [60] sont les deux algorithmes les plus importants et les plus largement utilisés dans un large éventail d'applications [6] [71]. PCA est une méthode non supervisée qui n'exige pas que les échantillons d'apprentissage soient étiquetés, tandis que LDA est une méthode supervisée qui utilise des informations spécifiques à la classe. Parmi les autres algorithmes classiques d'apprentissage de sous-espaces linéaires incluent ICA (Independent Component Analysis) [72], LPP (Locality Preserving Projection) [73] et CCA (Canonical Correlation Analysis) [74].

Ces algorithmes d'apprentissage de sous-espace linéaire sont des algorithmes opérant sur des vecteurs, c'est-à-dire des tenseurs du premier ordre. Pour appliquer ces méthodes d'apprentissage de sous-espace linéaire à des données de tenseurs d'ordre supérieur, telles que les images et les vidéos, nous devons d'abord reformer (vectoriser) les tenseurs en vecteurs, c'est-à-dire convertir les tableaux à N dimensions ($N > 1$) en tableaux à une dimension. Ainsi, l'apprentissage de sous-espace linéaire n'atténue que partiellement la malédiction de la dimensionnalité, tandis que cette vectorisation présente deux limitations fondamentales :

- La vectorisation casse la structure naturelle et la corrélation dans les données originales, réduit les redondances et / ou les dépendances d'ordre supérieur présentes dans les bases de données et perd des représentations potentiellement plus compactes ou utiles pouvant être obtenues sous les formes de tenseurs originales.
- Pour les données de tenseur d'ordre élevé de grande taille telles que les séquences vidéo, les vecteurs remodelés sont de très haute dimension. L'analyse de ces vecteurs entraîne souvent une demande de calcul et de mémoire importante (parfois irréalisable), ainsi que des difficultés liées à la taille réduite de l'échantillon en raison du grand nombre de paramètres à estimer.

2.6 Apprentissage de sous-espace multilinéaire

2.6.1 Définition du problème

L'apprentissage de sous-espace multilinéaire est l'extension multilinéaire de l'apprentissage de sous-espace linéaire. Il résout pour une projection multilinéaire avec quelques critères d'optimalité. Le problème impliqué dans l'approche d'apprentissage de sous-espace multilinéaire basé sur la projection du tenseur en tenseur peut être formulé mathématiquement d'une manière similaire à celle de la définition de l'apprentissage de sous-espace linéaire donnée à la section 2.5 comme suit :

Un ensemble d'échantillons tensoriels M d'ordre N est disponible pour l'apprentissage $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M\}$, où chaque échantillon \mathbf{A}_m est un tenseur $I_1 \times I_2 \times \dots \times I_N$ dans un espace tensoriel $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. L'objectif d'apprentissage de sous-espace multilinéaire est de trouver une transformation multilinéaire (matrices de projection) $Y^{(n)} \in \mathbb{R}^{I_n \times P_n}$ pour tous $n=1,2,\dots,N$. Ces matrices de transformation sont utilisées pour transformer ou mapper chaque tenseur de l'espace d'origine dans un espace d'ordre inférieur $\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \dots \otimes \mathbb{R}^{P_N}$ avec $P_n < I_n$ pour tous $n=1,2,\dots,N$, en utilisant la fonction de projection $\mathbf{B} = \mathbf{A} \times_1 Y^{(1)T} \times_2 Y^{(2)T} \times \dots \times_N Y^{(N)T}$ telle que les échantillons projetés (les caractéristiques extraites) satisfassent à un critère d'optimalité. Ce problème est résolu à l'aide d'une méthode de projection itérative. Le problème d'apprentissage de sous-espace multilinéaire peut être écrit mathématiquement comme suit :

$$\{Y^{(n)}, n = 1, 2, \dots, N\} = \arg \max_{U^{(1)}, U^{(2)}, \dots, U^{(N)}} \varphi(\mathbf{B}) \quad (2.30)$$

ou

$$\{Y^{(n)}, n = 1, 2, \dots, N\} = \arg \min_{U^{(1)}, U^{(2)}, \dots, U^{(N)}} \varphi(\mathbf{B}) \quad (2.31)$$

où $\varphi(\cdot)$ désigne une fonction de critère à maximiser ou à minimiser.

2.6.2 Mesures de dispersion pour les tenseurs

Par analogie avec la définition de divers dispersés pour des caractéristiques vectorielles dans un apprentissage linéaire en sous-espace dans [60], nous définissons les dispersions basées sur le tenseur dans un apprentissage de sous-multilinéaire par la projection du tenseur en tenseur.

Définition 2.1

Soit $\{\mathbf{A}_m, m = 1, \dots, M\}$ un ensemble de M échantillons de tenseur dans $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$. La dispersion totale de ces tenseurs est définie par:

$$\psi_{\mathbf{A}} = \sum_{m=1}^M \|\mathbf{A}_m - \bar{\mathbf{A}}\|_{\mathbb{F}}^2 \quad (2.32)$$

où $\bar{\mathbf{A}}$ est le tenseur moyen calculé comme

$$\bar{\mathbf{A}} = \frac{1}{M} \sum_{m=1}^M \mathbf{A}_m \quad (2.33)$$

La matrice de dispersion totale mode k de ces échantillons est ensuite définie comme suit:

$$G_{\mathbf{T}_A}^{(k)} = \sum_{m=1}^M (A_{m(k)} - \bar{A}_{(k)})(A_{m(k)} - \bar{A}_{(k)})^T \quad (2.34)$$

où $A_{m(k)}$ et $\bar{A}_{(k)}$ sont le déploiement en mode k de \mathbf{A}_m et $\bar{\mathbf{A}}$, respectivement.

Définition 2.2

Soit $\{\mathbf{A}_m, m = 1, \dots, M\}$ un ensemble de M échantillons de tenseur dans $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$. La dispersion inter-classe de ces tenseurs est définie par:

$$\psi_{B_A} = \sum_{c=1}^C M_c \|\bar{\mathbf{A}}_c - \bar{\mathbf{A}}\|_{\mathbb{F}}^2 \quad (2.35)$$

et la dispersion intra-classe de ces tenseurs est définie par

$$\psi_{W_A} = \sum_{c=1}^C \|\mathbf{A}_m - \bar{\mathbf{A}}_{c_m}\|_{\mathbb{F}}^2 \quad (2.36)$$

où C est le nombre de classes, M_c est le nombre d'échantillons pour la classe c , c_m est l'étiquette de la classe pour le $m^{\text{ième}}$ échantillon \mathbf{A}_m , $\bar{\mathbf{A}}$ est le tenseur moyen, et le tenseur moyen de la classe est

$$\bar{\mathbf{A}}_c = \frac{1}{M_c} \sum_{m, c_m=c} \mathbf{A}_m \quad (2.37)$$

Ensuite, les matrices de dispersion de mode k inter-classe et intra-classe sont définies en conséquence.

La matrice de dispersion inter-classe de mode k de ces échantillons est définie par

$$S_{BA}^{(k)} = \sum_{c=1}^C M_c (\bar{A}_{c(k)} - \bar{A}_{(k)}) (\bar{A}_{c(k)} - \bar{A}_{(k)})^T \quad (2.38)$$

et la matrice de dispersion intra-classe mode k de ces échantillons est définie par

$$S_{WA}^{(k)} = \sum_{m=1}^M M_c (A_{m(k)} - \bar{A}_{c_{m(k)}}) (A_{m(k)} - \bar{A}_{c_{m(k)}})^T \quad (2.39)$$

où $\bar{A}_{c_{m(k)}}$ est le mode k déploiement de \bar{A}_{c_m}

À partir des définitions ci-dessus, les propriétés suivantes sont dérivées.

Propriété 2.8 Puisque $\text{tr}(AA^T) = \|A\|_F^2$ et $\|A\|_F^2 = \|A_{(k)}\|_F^2$

$$\psi_{BA} = \text{tr}(S_{BA}^{(k)}) = \sum_{c=1}^C M_c \|\bar{A}_{c(k)} - \bar{A}_{(k)}\|_F^2 \quad (2.40)$$

Propriété 2.9

$$\psi_{WA} = \text{tr}(S_{WA}^{(k)}) = \sum_{m=1}^M M_c \|\bar{A}_{m(k)} - \bar{A}_{c_{m(k)}}\|_F^2 \quad (2.41)$$

2.7 Réduction de la dimensionnalité par l'apprentissage de sous-espace

Les données de tenseur du monde réel sont généralement spécifiées dans un espace de grande dimension. Le fonctionnement direct sur cet espace souffre de la malédiction de la dimensionnalité :

- Le traitement de données de grande dimension exige beaucoup de puissance de traitement et de ressources, de sorte que le calcul est coûteux [4].
- Lorsque le nombre d'échantillons de données disponibles est faible par rapport à leur grande dimensionnalité, c'est-à-dire que, dans le scénario de petite taille d'échantillon, les outils classiques deviennent inadéquats et de nombreux problèmes deviennent mal traités ou mal conditionnés [75].

La réduction de dimensionnalité est une tentative de transformer une base de données de grande dimension en une représentation de faible dimension tout en conservant la plupart des informations relatives à la structure sous-jacente [76]. En d'autres termes, c'est une projection d'un espace d'entrée de grande dimension vers un espace de sortie de petite dimension qui est un sous-espace de l'espace d'entrée. Contrairement à la projection linéaire, la projection multilinéaire ne transforme pas les tenseurs en vecteurs et l'estimation de paramètres pour chaque mode est ainsi obtenue de manière séquentielle et itérative jusqu'à la convergence. Ainsi, les bases de la projection multilinéaire sont plus contraintes, de sorte que les représentations apprises ont plus de régularités. Cela apporte trois avantages clés à l'apprentissage de sous-espace multilinéaire :

- Il préserve la structure des données avant la projection en prenant directement les tenseurs en entrée.
- Il peut apprendre des représentations plus compactes et potentiellement plus utiles que l'apprentissage linéaire de sous-espace. Avec la même quantité de données, l'apprentissage de sous-espace multilinéaire a moins de problèmes de petite taille d'échantillon que l'apprentissage de sous-espace linéaire.
- Il peut gérer les données de gros tenseurs plus efficacement avec des calculs des dimensions beaucoup plus basses que les méthodes linéaires.

La procédure complète d'apprentissage de sous-espace multilinéaire est fournie dans l'algorithme 2.1. Le tableau 2.2 résume les principales différences entre l'apprentissage de sous-espace multilinéaire et l'apprentissage de sous-espace linéaire. Dans le tableau 2.2, les données volumineuses font référence aux données dont la dimensionnalité dépasse la puissance de traitement du matériel informatique classique lorsque des algorithmes d'apprentissage de sous-espace linéaires sont utilisés, telles que des images de visage avec une résolution très élevée.

Algorithme 2.1: Pseudo-code typique d'apprentissage de sous-espace multilinéaire

Entrée: Un ensemble d'échantillons des tenseurs $\mathbf{A}_m \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, $m = 1, \dots, M$, et la dimensionnalité désirée du sous-espace tensoriel.

Sortie: La projection multilinéaire qui maximise un critère d'optimalité dans l'espace projeté.

Algorithme:

Etape 1 (Initialisation): Initialise la projection multilinéaire.

Etape 2 (Optimisation locale):

(a) **Pour** $n=1 : T$

➤ **Pour** $k=1 : N$

Résoudre pour la projection multilinéaire en mode k

➤ Si $n = T$ ou l'algorithme converge, Casser et sortir la projection multilinéaire actuelle.

Tableau 2.1: Comparaison entre l'apprentissage de sous-espace linéaire et multilinéaire.

Apprentissage de sous-espace linéaire	Apprentissage de sous-espace multilinéaire
Remodeler en vecteurs	Travailler directement sur la représentation tensorielle naturelle
Casser la structure naturelle	Préserver la structure naturelle
Estimer un grand nombre de paramètres	Estimer moins de paramètres
Problème plus sévère de petite taille d'échantillon	Problème moins sévère de petite taille d'échantillon
Peu applicable aux données massives	Capable de gérer des données massives
Solution fermée	Solution sous-optimale et itérative

2.8 Conclusion

Dans ce chapitre, nous avons introduit les principes fondamentaux à la bonne compréhension d'algèbre multilinéaire. Ainsi, nous avons pu définir la notion de tenseur qui permet de travailler directement avec les données sous forme de tableaux multidimensionnels. Nous avons ensuite présenté les trois types de base de la projection multilinéaire : la projection vecteur en vecteur, la projection tenseur en tenseur et la projection tenseur en vecteur. De plus, nous avons révélé les liens entre ces trois projections. Enfin, nous avons terminé ce chapitre par une présentation du problème d'apprentissage de sous-espace linéaire et multilinéaire. Dans le chapitre suivant, nous présentons un état de l'art des approches d'apprentissage de sous-espace basé sur l'analyse multidimensionnelle

Chapitre 3

Etat de l'art sur l'analyse multidimensionnelle pour la reconnaissance faciale

3.1 Introduction	43
3.2 Historique d'apprentissage de sous-espaces multilinéaire	43
3.3 Revue sur les algorithmes d'apprentissage de sous-espaces multilinéaires pour la reconnaissance faciale	44
3.3.1 Approches d'apprentissage de sous-espaces multilinéaires non supervisées	46
3.3.1.1 Techniques d'apprentissage de sous-espaces multilinéaires non supervisées par la projection tenseur en tenseur	46
3.3.1.2 Techniques d'apprentissage de sous-espace multilinéaire non supervisées par la projection tenseur en vecteur	57
3.3.3 Approches d'apprentissage de sous-espaces multilinéaires supervisées	59
3.3.3.1 Techniques d'apprentissage de sous-espace multilinéaires supervisées par la projection tenseur en tenseur	60
3.3.3.2 Techniques d'apprentissage de sous-espace multilinéaires supervisées par la projection tenseur en vecteur	64
3.4 Conclusion.....	68

3.1 Introduction

Les images de visage se présentent naturellement sous la forme de tenseurs d'ordre 2 ou d'ordre supérieur N , dans les domaines les plus actifs de la recherche biométrique et en particulier dans la reconnaissance faciale, l'utilisation de tenseurs est devenue une direction de recherche très importante [54, 77]. L'un des problèmes de l'analyse des données en reconnaissance de formes et en vision par ordinateur consiste à trouver un type de représentation de données approprié [78]. Les approches d'apprentissage de sous-espaces multilinéaire agissent directement sur les données tensorielles naturelles, ce qui résout le problème de la représentation tensoriel dans l'extraction de caractéristiques et la classification [12, 23].

Ce chapitre donne un aperçu des techniques d'apprentissage de sous-espaces multilinéaire pour la reconnaissance faciale, permettant de réduire la dimensionnalité des données multidimensionnelles à partir de leurs représentations originales. Nous commençons ce chapitre par une présentation de l'historique d'apprentissage de sous-espaces multilinéaire. Par la suite, nous présentons la littérature d'algorithmes existants de sous-espaces multilinéaire selon les deux classes supervisés et non supervisés.

3.2 Historique d'apprentissage de sous-espaces multilinéaire

L'apprentissage de sous-espaces multilinéaire a été bien étudié en mathématiques au milieu du XXe siècle. Il s'appuie sur le concept de tenseurs et développe la théorie des espaces tensoriels [66, 79, 80]. Avec la considération des données comme des entités indivisibles, de nouveaux outils mathématiques ont été développés. Ceux-ci reposent sur l'algèbre multilinéaire. Ainsi, depuis une quarantaine d'années, les décompositions tensorielles sont étudiées, et ce pour divers champs d'applications tels que la psychométrie, la chimiométrie, le traitement du signal et des images [81-84],

Les décompositions tensorielles ont été introduites par Hitchcock en 1927 [85], mais la considération d'un modèle multidimensionnel est attribuée à Catell en 1944 [86, 87]. De tels modèles n'ont guère attiré l'attention des chercheurs jusque dans les années 60 avec Tucker [84], Carroll et Chang [81], ainsi que Harschman [88] en 1970. Récemment, un livre sur l'analyse de données par des méthodes tensorielles est parue [89] ainsi que des travaux dans des domaines variés. Ces outils de décompositions tensorielles ont permis de généraliser aux tenseurs la SVD (singular value decomposition). Cependant, cette généralisation n'est pas immédiate du fait de la complexité de la modélisation tensorielle. Dans le cas matriciel, la SVD

peut s'interpréter comme étant un produit matriciel ou une somme de K matrices de rang 1 de façon équivalente. Dans le cas de tenseurs, cette équivalence n'est plus vérifiée, mais conduit à deux types de décompositions tensorielles. Le premier modèle est la décomposition tensorielle TUCKER3, développée par Tucker [84]. Elle permet de définir la supérieurs HOSVD (Higher-Order Singular Value Decomposition) [84] ainsi que l'approximation de rang inférieur-(K_1, K_2, \dots, K_N) d'un tenseur d'ordre N . Le second modèle généralise le concept de décomposition canonique et communément connu sous le nom de décomposition PARAFAC. Cette décomposition permet d'écrire un tenseur de rang R sous sa forme canonique, c'est-à-dire une somme de R tenseurs de rang 1.

Les travaux dans [47, 48] ont conduit de développement de nouveaux algorithmes multilinéaires et l'exploration de nouveaux domaines d'application pour l'analyse de données tensorielles. L'analyse multilinéaire de données d'image est initiée par la méthode TensorFace [90, 91], qui utilise les algorithmes multilinéaires proposés dans [47, 48] pour analyser les facteurs impliqués dans la formation d'images faciales. Cependant, dans ces travaux, les données d'entrée telles que des images ou des séquences vidéo sont toujours représentées sous forme de vecteurs. Ces vecteurs sont disposés en un tenseur en fonction de multiples facteurs impliqués dans leur formation en vue d'une analyse ultérieure. Une telle formation de tenseurs nécessite un grand nombre d'échantillons d'apprentissage capturés dans diverses conditions, ce qui est souvent peu pratique et peut poser le problème de données manquantes. De plus, la taille des données du tenseur est généralement énorme, ce qui entraîne un grand espace mémoire et des exigences de calcul élevées. Ces dernières années, plusieurs méthodes ont été proposées pour l'apprentissage direct d'un sous-espace à partir de données tensorielles [12, 17, 35, 64, 92-96].

3.3 Revue sur les algorithmes d'apprentissage de sous-espaces multilinéaires pour la reconnaissance faciale

Comme mentionné précédemment dans la section 2.5, les méthodes d'apprentissage de sous-espaces linéaire telles que PCA, LDA, NPP et LPP dans la reconnaissance du visage nécessite de réorganiser les échantillons de données en longs vecteurs dans un espace de très grande dimension. Cela nécessite un grand nombre de paramètres à estimer, ainsi entraîne des coûts de calcul et des exigences de mémoire élevés. Au-delà des problèmes de mise en œuvre, la réorganisation de l'échantillon de données casse la structure et la corrélation naturelles et

supprime les redondances et les dépendances dans leurs formes tensorielles naturelles dans laquelle des représentations plus compactes peuvent être obtenues [18, 92, 94]. Ceci a inspiré le développement récent d'algorithmes d'apprentissage de sous-espaces multilinéaire [12, 24, 97, 98] opérant directement sur les représentations tensorielles que sur leurs formes vectorielles. Jusqu'à maintenant, de nombreuses techniques d'apprentissage de sous-espaces multilinéaire ont été proposées pour la reconnaissance de visage basé sur l'analyse multidimensionnel de données, elles sont divisées en deux catégories multilinéaires principales : les approches non supervisées et les approches supervisée. Dans les paragraphes suivants, nous présentons chaque catégorie avec ses algorithmes existants dans les travaux précédents.

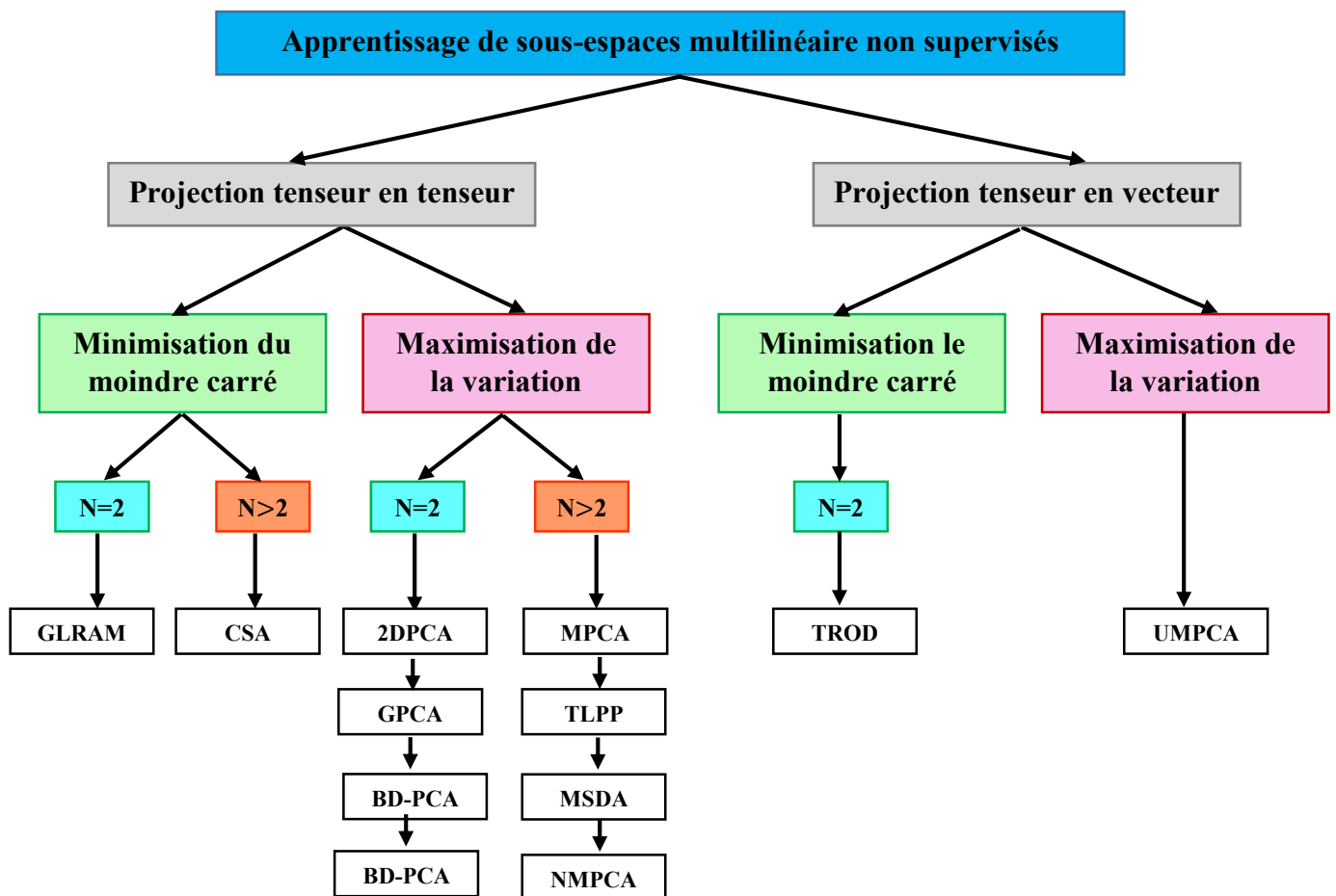


Figure 3.1: Taxonomie des algorithmes d'apprentissage de sous-espaces multilinéaires non supervisé.

3.3.1 Approches d'apprentissage de sous-espaces multilinéaires non supervisées

Les approches d'apprentissage non supervisé prennent des données sans étiquettes pour l'apprentissage, en essayant de trouver une structure cachée. Il peut être utilisé pour la classification ou la régression avec un classifieur ou régresseur. Le développement d'algorithmes non supervisé a commencé avec le traitement d'images directement sous forme de matrices plutôt que de vecteurs. La figure 3.1 donne un aperçu sur les algorithmes d'apprentissage de sous-espaces multilinéaire non supervisés.

Nous pouvons distinguer deux types de techniques parmi les approches multilinéaire non supervisés: les techniques du projection tenseur en tenseur et les techniques du projection tenseur en vecteur.

3.3.1.1 Techniques d'apprentissage de sous-espaces multilinéaires non supervisées par la projection tenseur en tenseur

- **2DPCA (Two-Dimensional Principal Component Analysis)**

Yang et al [92] ont proposé une méthode non supervisé pour la représentation d'image appelée 2DPCA (Two-Dimensional Principal Component Analysis). Contrairement à PCA, 2DPCA est basé sur des matrices d'images 2D au lieu sur des vecteurs 1D, donc Il n'est pas nécessaire de transformer la matrice d'images en un vecteur avant l'extraction des caractéristiques. Cet algorithme cherche à trouver une transformation linéaire $Y \in \mathbb{R}^{I_2 \times P_2}$ ($P_2 < I_2$) qui projette une image $A_k \in \mathbb{R}^{I_1 \times I_2}$ ($K=1,2,\dots, M$) vers une matrice de faible dimension V_k

$$V_k = A_k Y = A_k \times_2 Y^T \in \mathbb{R}^{I_1 \times P_2} \quad (3.1)$$

Plus précisément, 2DPCA essaie de maximiser la mesure de dispersion suivante :

$$\sum_{k=1}^M \|V_k - \bar{V}\|_F^2 = \sum_{k=1}^M \text{trace}(Y^T S_x Y) \quad (3.2)$$

$$\sum_{k=1}^M \|V_k - \bar{V}\|_F^2 = \sum_{k=1}^M \text{trace}(Y^T (A_k - \bar{A})^T (A_k - \bar{A}) Y) \quad (3.3)$$

Où \bar{A} est l'image moyenne $\bar{A} = \frac{1}{M} \sum_{K=1}^M A_k$.

Ainsi, la matrice de projection optimale $S_{\text{opt}} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d]$ peut être obtenue en évaluant les vecteurs propres correspondant aux d premières valeurs propres les plus grandes de S . Les données d'image sont projetées dans le mode 2 (mode ligne) uniquement, tandis que la

projection dans le mode 1 (mode colonne) est ignorée, ce qui entraîne une mauvaise performance. La projection dans cette technique est considérée comme un cas particulier de la projection tenseur en tenseur du second ordre, où la matrice de projection à mode 1 est une matrice d'identité.

Evidemment, les résultats expérimentaux dans [92] sur plusieurs bases de données de visage : (ORL, AR et Yale) ont montré l'amélioration et l'efficacité de l'extraction de caractéristiques avec 2DPCA par rapport à la méthode classique PCA.

- **GPCA (Generalized Principal Component Analysis)**

En revanche, Ye et al [17] ont proposé une nouvelle stratégie de réduction des dimensions, appelée GPCA (Generalized Principal Component Analysis), cette méthode est une extension de PCA pour les données matricielles, c'est-à-dire les tenseurs du second ordre. C'est un algorithme non supervisé visant à maximiser la variation capturée dans le sous-espace projeté comme dans PCA classique. Tandis que PCA utilise une représentation vectorielle de la matrice d'image 2D, GPCA utilise une représentation matricielle bidimensionnelles 2D (comme illustré schématiquement à la figure 3.2) et fonctionne directement avec les images dans leur état originales.

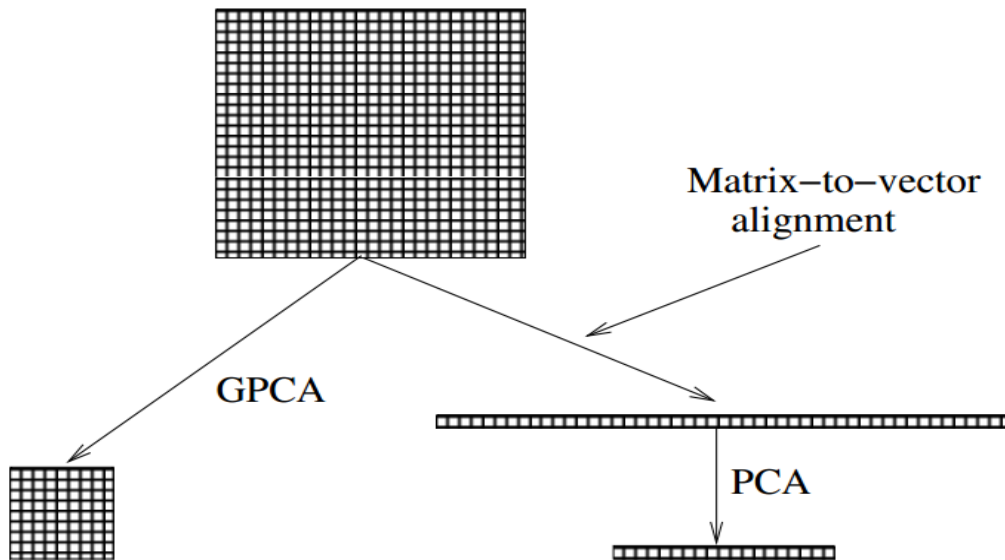


Figure 3.2: Vue schématique de la différence essentielle entre GPCA et PCA [17].

Comme dans PCA, l'algorithme GPCA prend les données centrées comme entrée $\check{A}_K = A_K - \bar{A}$ plutôt que les coordonnées d'origine A_K . Le problème GPCA est défini comme suit :

un ensemble de M échantillons de données matricielles $\{A_1, A_2, \dots, A_M\}$. Chaque échantillon $\{A_k \in \mathbb{R}^{I_1 \times I_2}, k = 1, \dots, M\}$ assume des valeurs dans l'espace tenseur $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2}$, qui est défini comme le produit tensoriel de deux espaces vectoriels \mathbb{R}^{I_1} et \mathbb{R}^{I_2} . L'objectif de GPCA est de trouver deux transformations linéaires (deux matrices de projection) $\check{Y}^{(1)} \in \mathbb{R}^{I_1 \times P_1}$ ($P_1 \leq I_1$) et $\check{Y}^{(2)} \in \mathbb{R}^{I_2 \times P_2}$ ($P_2 \leq I_2$), avec colonnes ortho-normales, qui projettent un échantillon A_k à

$$\check{V}_K = \check{Y}^{(1)T} \check{A}_K \check{Y}^{(2)} = \check{A}_K \times_1 \check{Y}^{(1)T} \times_2 \check{Y}^{(2)T} \in \mathbb{R}^{P_1 \times P_2} \quad (3.4)$$

Tel que $\{\check{Y}_K \in \mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2}, K = 1, \dots, M\}$ capture la majeure partie de la variation observée dans les échantillons d'origine, originaux, en supposant que la variation est mesurée par la dispersion totale Ψ , la fonction objective pour GPCA définit comme suit :

$$\begin{aligned} (\check{Y}^{(1)}, \check{Y}^{(2)}) &= \arg \max_{Y^{(1)}, Y^{(2)}} \Psi = \arg \max_{Y^{(1)}, Y^{(2)}} \sum_{K=1}^M \|V_k - \bar{V}\|_F^2 \\ &= \arg \max_{Y^{(1)}, Y^{(2)}} \sum_{K=1}^M \|Y^{(1)T} \check{A}_K Y^{(2)}\|_F^2 \end{aligned} \quad (3.5)$$

Des expériences sur des bases de données d'images de visages : (PIX, ORL, AR et Yale B) montrent que, pour le même espace de stockage, GPCA est supérieur à PCA en termes de qualité des images compressées, de précision et de coût de calcul. La figure 3.3 représente les résultats de la projection des données avec GPCA en utilisant des coordonnées homogènes.

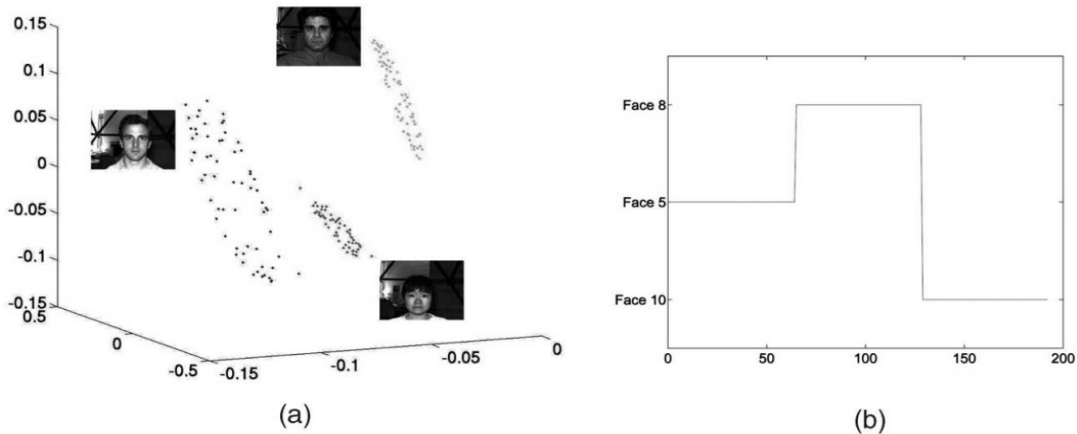


Figure 3.3: Regroupement d'un sous-ensemble de la base de données de visages Yale B sous des conditions d'illumination pour les sujets 2, 5 et 8. (a) Données d'image projetées sur les trois composants principaux. (b) Résultats de regroupement par la méthode GPCA [99].

- **BD-PCA (Bidirectional Principal Component Analysis)**

Zuo et al [100] ont proposé BD-PCA (Bidirectional Principal Component Analysis) pour traiter simultanément deux problèmes : l'extraction de caractéristiques et la classification de visages. La projection avec cette méthode est considérée comme un cas particulier de la projection tenseur à tenseur du second ordre. BD-PCA directement exécuté sur la matrice d'image originale en réduisant la dimensionnalité dans les directions des colonnes et des lignes dans laquelle une matrice de caractéristiques V ($K_{col} \times K_{lig}$) d'une image A ($m \times n$) où ($K_{col} \ll m, K_{lig} \ll n$) peut être obtenu par

$$V = Y_{col}^T A Y_{lig} \quad (3.6)$$

Où Y_{col} est le projecteur de colonne et Y_{lig} est le projecteur de ligne. BD-PCA présente au moins deux avantages par rapport à PCA classique. Tout d'abord, étant une technique de projection d'image simple, BD-PCA ne nécessite pas de rechape (transforme) d'une image A sur un vecteur d'image a . Deuxièmement, l'extraction de caractéristiques basée sur BD-PCA nécessite généralement peu de calcul que l'extraction de caractéristiques basée sur PCA.

Soit $\{A_1, A_2, \dots, A_M\}$ être un ensemble d'apprentissage de M images. En représentant la matrice d'image A_i en tant que m -ensemble de $1 \times n$ lignes de vecteurs, la matrice de dispersion totale de ligne S_t^{lig} est définie par

$$S_t^{lig} = \frac{1}{Nm} \sum_{i=1}^N (A_i - \bar{A})^T (A_i - \bar{A}) \quad (3.7)$$

Où \bar{A} est la matrice moyenne de toutes les images d'apprentissage. La matrice de projection de ligne est construite en choisissant les vecteurs propres de lignes correspondant aux K_{lig} premières valeurs propres les plus grandes de S_t^{lig} .

$$Y_{lig} = [Y_1^{lig}, Y_2^{lig}, \dots, Y_{K_{lig}}^{lig}] \quad (3.8)$$

En traitant une matrice d'image A_i comme un n ensemble de $m \times 1$ vecteurs de colonne, la matrice de dispersion totale de colonne S_t^{col} est définie par

$$S_t^{col} = \frac{1}{Nm} \sum_{i=1}^N (A_i - \bar{A})^T (A_i - \bar{A}) \quad (3.9)$$

La matrice de projection de colonne est construite en choisissant les vecteurs propres de colonnes correspondant aux K_{lig} premières valeurs propres les plus grandes de S_c^{col} .

$$Y_{lig} = [Y_1^{lig}, Y_2^{lig}, \dots, Y_{K_{lig}}^{lig}] \quad (3.10)$$

Notez que BD-PCA est une généralisation du 2DPCA de Yang [92]. 2DPCA peut être considéré comme un BD-PCA spécial avec $Y_{col} = I_m$, où I_m , désigne une matrice d'identité $m \times m$.

Des expériences approfondies dans [100] ont été réalisées à l'aide de l'algorithme de reconnaissance faciale BD-PCA ont permis d'obtenir des performances supérieures par rapport aux techniques de l'état de l'art. La méthode BD-PCA a été améliorée dans [101].

- **GLRAM (Generalized Low Rank Approximation of Matrices)**

Ye [18] a développé un algorithme plus général appelée GLRAM (Generalized Low Rank Approximation of Matrices). Il prend en compte la spatiale corrélation des pixels de l'image dans un voisinage localisé et applique deux transformations linéaires sur les côtés gauches et les côtés droits de matrices d'images d'entrées. Cet algorithme résout deux transformations linéaires qui projettent une image sur une matrice de faible dimension tout en minimisant la mesure d'erreur du moindre carré (reconstruction). Ainsi, des projections dans les deux modes sont impliquées et la projection effectuée est tenseur en tenseur du second ordre ($N = 2$). Comme GLRAM et 2DPCA fonctionne avec des données sous forme de matrice, la différence principale est que 2DPCA applique une transformation linéaire sur le côté droit des données, tandis que GLRAM applique une transformation linéaire à deux côtés. Cet algorithme résout deux transformations linéaires $Y^{(1)} \in \mathbb{R}^{I_1 \times P_1}$ ($P_1 < I_1$) et $Y^{(2)} \in \mathbb{R}^{P_2 \times I_2}$ ($P_2 < I_2$) qui projettent une image $A_k \in \mathbb{R}^{I_1 \times I_2}$ à une image de faible dimension V_k

$$V_k = Y^{(1)} A_k Y^{(2)} = A_k \times_1 Y^{(1)T} \times_2 Y^{(2)T} \in \mathbb{R}^{P_1 \times P_2} \quad (3.11)$$

En minimisant la mesure d'erreur des moindres carrés

$$\sum_{k=1}^M \| A_k - Y^{(1)} V_k Y^{(2)T} \|_F^2 = \sum_{k=1}^M \| A_k - V_k \times_1 Y^{(1)} \times_2 Y^{(2)} \|_F^2 \quad (3.12)$$

des meilleurs résultats de réduction de la dimensionnalité sont obtenus selon [18] que dans [92]. La figure 3.4 compare les différents schémas de projection utilisés par l'approche GLRAM et l'approche proposé par Wang et Ahuja [102].

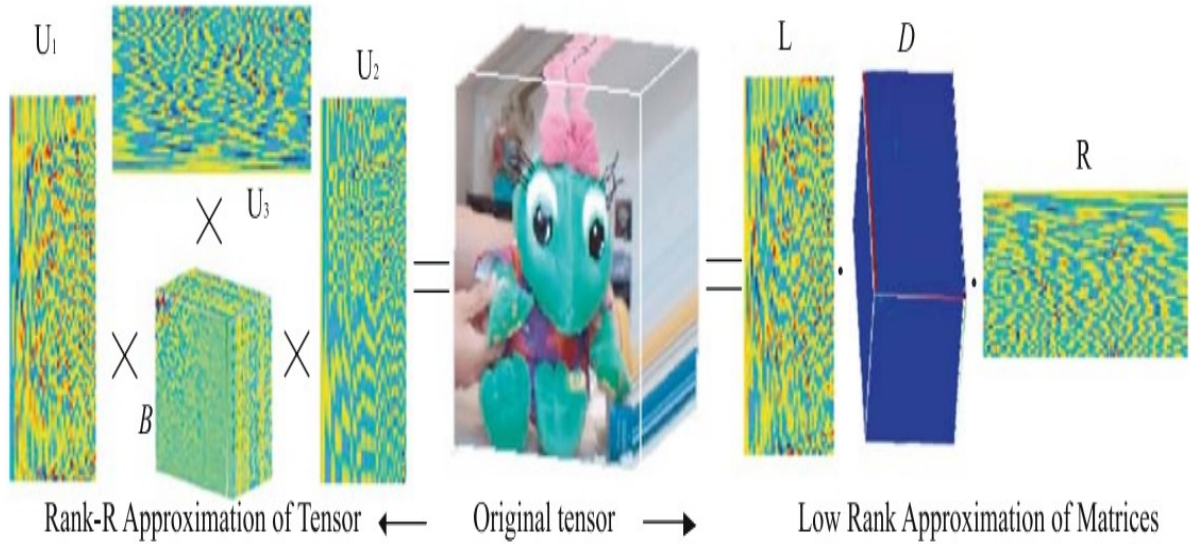


Figure 3.4: Illustration de la méthode Rank-R Approximations of Tensors vis-à-vis la méthode GLRAM. [102].

L'évaluation de la méthode GLRAM dans [18] sur plusieurs bases de données de visage : (PIX, ORL, AR, PIE et USPS) a montré les excellentes performances de réduction de dimensionnalité que 2DPCA.

- **CSA (Concurrent Subspaces Analysis)**

Bien que 2DPCA, BD-PCA et GLRAM exploitent les deux modes pour l'apprentissage en sous-espace, et formulées que pour les matrices (tenseurs d'ordre 2). Xu et al [35] ont introduit une méthode d'approximation tensorielle appelée CSA (Concurrent Subspaces Analysis) en tant que généralisation de GLRAM pour les tenseurs d'ordre supérieur. Cette méthode travaille directement sur la forme multidimensionnelle des données représentées sous forme de matrice en 2D et tenseur dans les dimensions supérieures comme illustré sur la figure 3.5.

CSA est un algorithme d'apprentissage de sous-espace multilinéaire non supervisé. Il résout pour une projection tenseur en tenseur $\{Y^{(n)} \in \mathbb{R}^{I_n \times P_n}, P_n \leq I_n, n = 1, \dots, N\}$ qui projette un tenseur $\mathbf{A}_m \in \mathbb{R}^{I_1 \times \dots \times I_N}$ vers un tenseur réduit \mathbf{V}_m .

$$\mathbf{V}_m = \mathbf{A}_m \times_1 Y^{(1)T} \times_2 \dots \times_N Y^{(N)T} \in \mathbb{R}^{P_1 \times \dots \times P_N} \quad (3.13)$$

En minimisant la métrique d'erreur de reconstruction suivante

$$\sum_{m=1}^M \|A_m - V_m \times_1 Y^{(1)} \times_2 \dots \times_N Y^{(N)}\|_F^2 \quad (3.14)$$

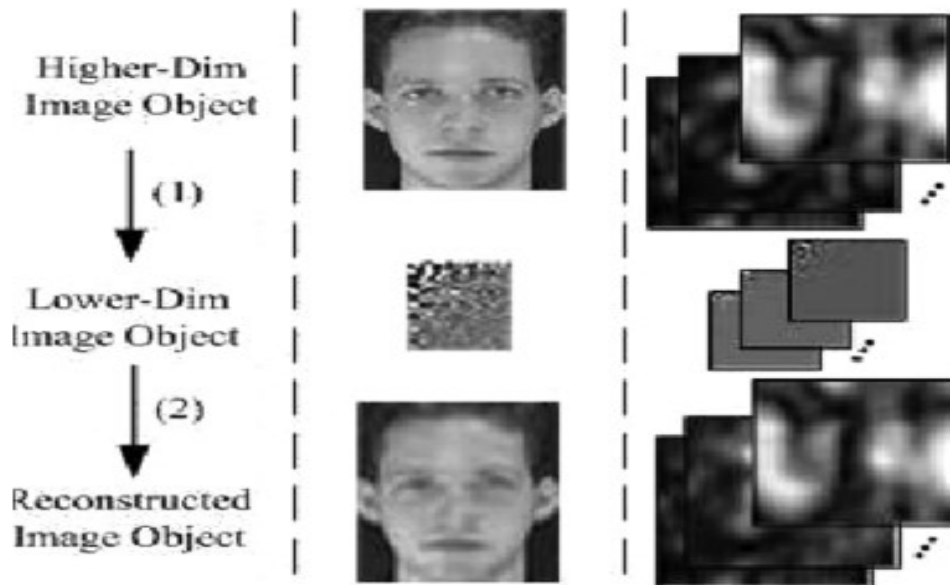


Figure 3.5: Réduction de la dimensionnalité et de la reconstruction des objets pour les matrices et les tenseurs d'ordre supérieur. (1) indique une réduction de la dimensionnalité, (2) reconstruction d'objet par l'algorithme CSA [35].

Les deux méthodes GLRAM et CSA sont formulés dans l'objectif de reconstruction optimale ou d'approximation des tenseurs. Par conséquent, ils ont ignoré une étape importante de centrage dans les algorithmes d'apprentissage en sous-espace non supervisés développés pour la reconnaissance, tels que le PCA classique, où les données sont centrées d'abord avant d'obtenir la projection en sous-espace.

Dans les applications de reconnaissance où les solutions impliquent des problèmes propres, le non-centrage (autrement dit, une moyenne différente de zéro) peut potentiellement affecter la décomposition propre dans chaque mode et conduire à une solution qui capture plutôt la variation par rapport à l'origine que de capturer la véritable variation des données (par rapport au centre de données) [24].

Des expériences approfondies dans [35] sur les bases de données de visage FERET, ORL et CMU PIE codés en tenseurs de second ou troisième ordre montrent que le CSA proposé surpasse les algorithmes linéaires basés sur la PCA en reconstruction et en reconnaissance de visage.

- **MPCA (Multilinear principal Component Analysis)**

Plus récemment, Lu et al [24] ont proposé la méthode MPCA (Multilinear principal Component Analysis) qui développé sur la base de l'algorithme PCA pour l'extraction et la classification des caractéristiques d'un objet tenseur. Son avantage est de fonctionné directement sur les formes tensorielles, contrairement aux algorithmes traditionnels qui transforment les données de grande dimension en un vecteur unidimensionnel. MPCA cherche dans chaque mode les bases permettant de projeter le tenseur d'entrés et de capturé la plupart des variations (les informations globales). La figure 3.6 montre la représentation du système de reconnaissance d'objets tensoriel basé sur la méthode MPCA.

L'objectif de la MPCA est de définir N matrices de projection $\{Y^{(n)} \in \mathbb{R}^{I_n \times P_n}, n = 1, \dots, N\}$ qui mappe l'espace tenseur d'origine $\mathbf{A}_M \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ dans un sous-espace tenseur réduit $\mathbf{V}_M \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ (avec $P_k < I_k$ pour $k = 1 \dots N$)

$$\mathbf{V}_M = \mathbf{A}_M \times_1 Y^{(1)T} \times_1 Y^{(2)T} \dots \times_N Y^{(N)T} \quad (3.15)$$

Les matrices de projection $\{Y^{(n)} \in \mathbb{R}^{I_n \times P_n}, n = 1, \dots, N\}$ peuvent être obtenues en maximisant la dispersion totale du déploiement les tenseurs en mode k. Le problème d'optimisation peut être représenté par :

$$Y^k = \operatorname{argmax} \operatorname{tr}((Y^k)^T T^k Y^k) \quad (3.16)$$

$$T^k = \sum_{i=1}^M (A_i^{(k)} - \bar{A}^{(k)})(A_i^{(k)} - \bar{A}^{(k)})^T \quad (3.17)$$

Où $A_i^{(k)}$ est la matrice dépliée en mode k de \mathbf{A}_i .

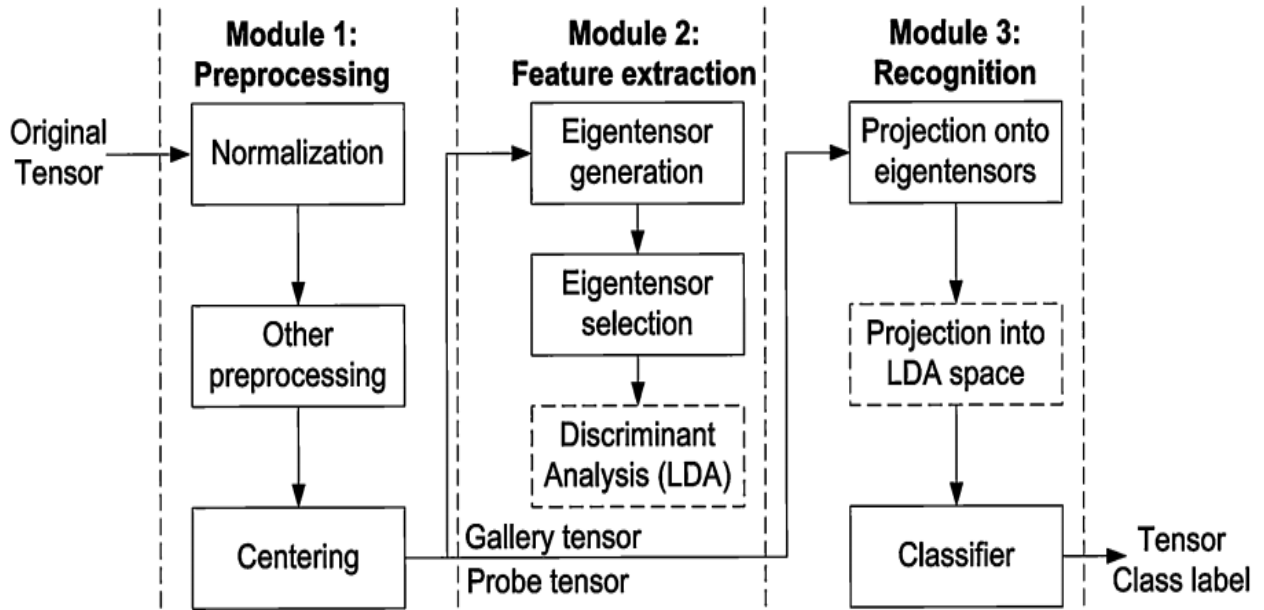


Figure 3.6: Schéma fonctionnel du système de reconnaissance d'objets tensoriel basé sur la méthode MPCA [24].

Les résultats expérimentaux dans [93] et [103] menés sur les bases de données de visage AT & T et FERET montrent que la méthode proposée (MPCA) convient mieux sur les grands ensembles de données, et permet d'obtenir une meilleure efficacité de calcul avec la supériorité de cet algorithme par rapport au PCA, 2D PCA, GLRAM.

- **TLPP (Tensor Locality Preserving Projection)**

Dai et Yeung [104] ont proposé le tenseur LPP (Tensor Locality Preserving Projection) comme une extension de la méthode basique LPP aux tenseurs d'ordre élevé. Cette méthode permet de caractériser efficacement les relations entre les dimensions d'une représentation tenseur. De plus, le TLPP permet également d'estimer naturellement les propriétés géométriques et topologiques locales intrinsèques de la variété incorporée dans un espace tensoriel. Ainsi, elle ne souffre pas de la malédiction de la dimensionnalité et de la forte demande informatique. L'algorithme TLPP vise à trouver M matrices de transformation $Y^k \in \mathbb{R}^{I_k \times \hat{I}_k}$ ($I_k > \hat{I}_k, K=1, \dots, M$) utilisées pour projeter chaque tenseur $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M}$ en optimisant la formulation suivante

$$\begin{aligned} & \operatorname{argmin} (Y^1, \dots, Y^M) \\ & = \sum_{i,j} \|\mathbf{A}_i \times_1 Y^1 \dots \times_M Y^M - \mathbf{A}_j \times_1 Y^1 \dots \times_M Y^M\|_F^2 S_{i,j} \end{aligned} \quad (3.18)$$

Où $S_{i,j}$ est la matrice de poids correspondante.

Le problème d'optimisation de cet algorithme peut être résolu en utilisant un schéma itératif utilisé dans [104, 105]. Le TLPP considère uniquement la structure locale en préservant la relation entre le voisinage des échantillons et l'ensemble de données avant et après la projection.

Des expériences de reconnaissance faciale dans [104] et [106] basées sur les bases de données de visage FERET et l'ORL montrent que le TLPP donne des résultats très impressionnants par rapport aux autres algorithmes linéaire : PCA, LDA et LPP

- **(MSDA) Multilinear Spatial Discriminant Analysis**

Plus récemment, Yuan et al [107] ont proposé une nouvelle technique de projection multilinéaire appelée MSDA (Multilinear Spatial Discriminant Analysis). Différente de la méthode MPCA qui vise à préserver la structure globale et la méthode TLPP qui est en faveur de la préservation de la structure locale, MSDA cherche un compromis entre la structure non locale (globale) et la structure locale afin de diriger ses informations discriminantes à partir de la plage de la structure non locale, et de la plage de la structure locale. Cette caractéristique de discriminant spatial confère au MSDA une capacité de préservation l'information utile plus puissante que la TLPP et la MPCA. La figure 3.7 présente les directions de projection pour les méthodes MPCA, TLPP, et MSDA.

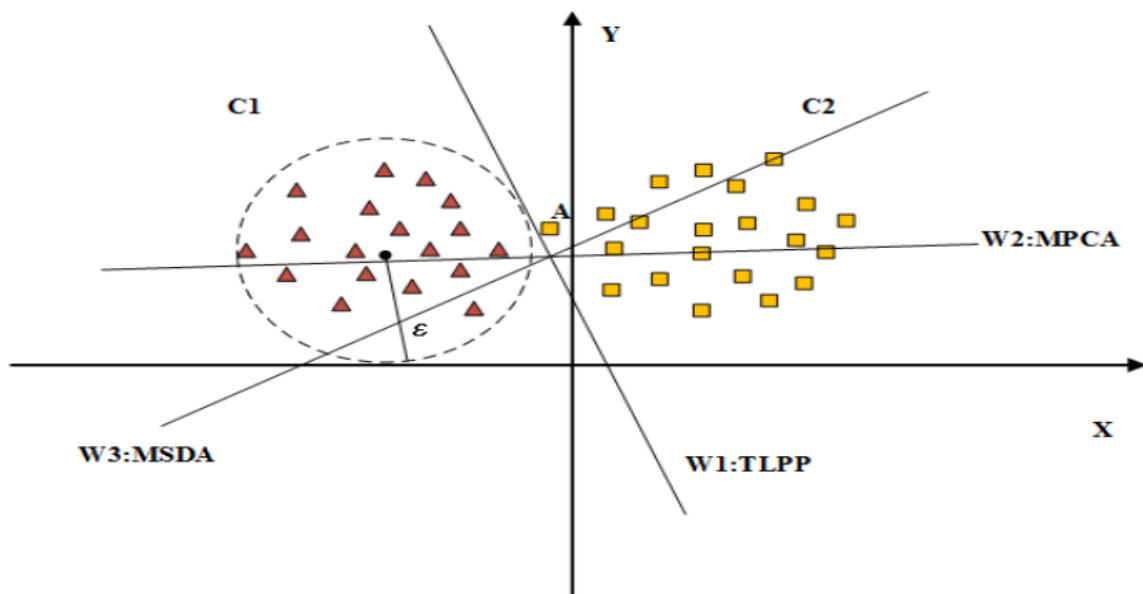


Figure 3.7: Illustration d'échantillons à deux classes dans un espace à deux dimensions et les directions de projection avec MPCA, TLPP, et MSDA [107].

L'objectif de la MSDA est de trouver M matrices de projection $Y^k \in \mathbb{R}^{I_k \times \hat{I}_k} (I_k > \hat{I}_k, K = 1, \dots, M)$ qui mappent les N modes tenseurs originaux dans un plus petit espace tenseur $\mathbb{R}^{\hat{I}_1 \times \hat{I}_2 \times \dots \times \hat{I}_M} (\hat{I}_1 \times \hat{I}_2 \times \dots \times \hat{I}_M \ll I_1 \times I_2 \times \dots \times I_M)$. La fonction objective de MSDA écrit comme suit

$$Y^k = \arg \max_{Y^k} \{ \text{Tr}(Y^{kT} W_N Y^k) - \eta \text{Tr}(Y^{kT} W_L Y^k) \} \quad (3.19)$$

Où $\eta (\eta > 0)$ est un paramètre permettant d'équilibrer la non-localité et la localité, W_N et W_L sont définies comme suit :

$$W_N = \sum_{i,j} (A_i^{(k)} - A_j^{(k)})(A_i^{(k)} - A_j^{(k)})^T S_{ij}^N \quad (3.20)$$

$$W_L = \sum_{i,j} (A_i^{(k)} - A_j^{(k)})(A_i^{(k)} - A_j^{(k)})^T S_{ij}^L \quad (3.21)$$

Où S_{ij}^L et S_{ij}^N sont les matrices de similarité locale et non local.

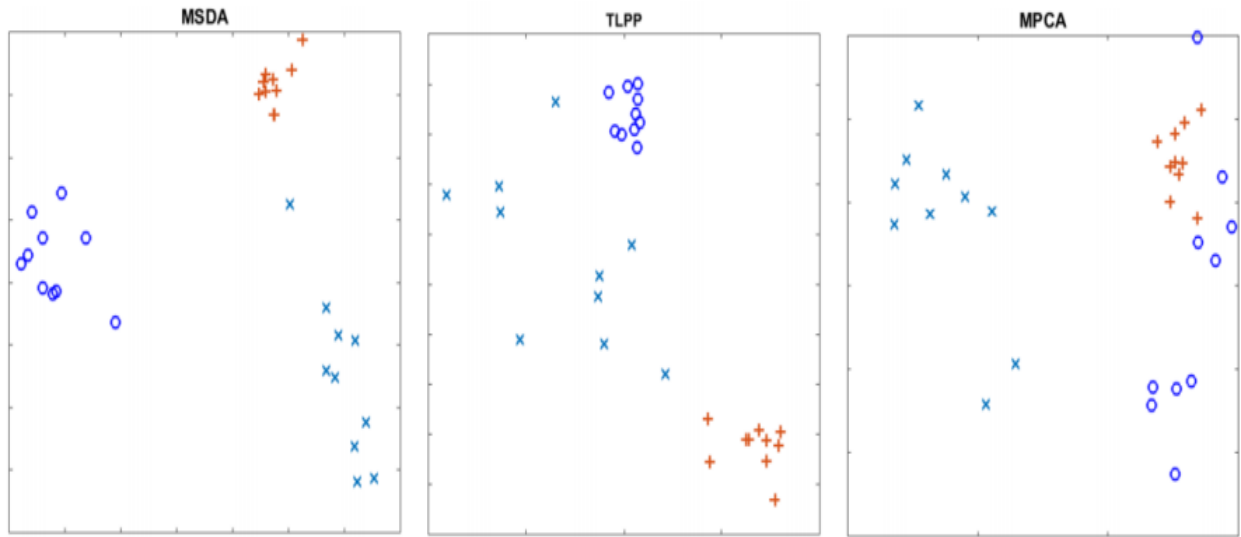


Figure 3.8: Résultats de la visualisation de différentes méthodes sur la base de données ORL [107].

Des expériences approfondies dans [107] sur des bases de données de visage ORL, CMU PIE et Yale-B montrent l'efficacité de la méthode proposée MSDA en terme de la séparation (classification) des données que d'autres méthodes multilinéaires TLPP et MPCA (voir figure 3.8).

3.3.1.2 Techniques d'apprentissage de sous-espaces multilinéaires non supervisées par la projection tenseur en vecteur

- **TROD (Tensor Rank-One Decomposition)**

L'algorithme TROD (Tensor Rank-One Decomposition) [25] introduit par Shashua et Levin est un algorithme non supervisé basé sur la projection tenseur en vecteur. Il est formulé pour les matrices d'images. Cet algorithme recherche une projection tenseur en vecteur du second ordre qui projette une image sur un vecteur de petite dimension tout en minimisant une mesure d'erreur par le moindre carré. Bien qu'il ait été formulé uniquement pour les matrices d'image, il peut être facilement étendu aux tenseurs d'ordre supérieur.

Donné M échantillons de tenseur d'apprentissage $\{\mathbf{X}_m\}$ le $P^{\text{ième}}$ projection multilinéaire élémentaire, noté comme $\{\mathbf{u}_p^{(n)}, n = 1, \dots, N\}$ projette le $m^{\text{ième}}$ échantillon \mathbf{X}_m sur un scalaire v_{m_p} comme suit

$$y_{m_p} = \mathbf{X}_m \times_1 \mathbf{z}_p^{(1)T} \times_2 \mathbf{z}_p^{(2)T} \dots \times_N \mathbf{z}_p^{(N)T} = \mathbf{X}_m \times_{n=1}^N \{\mathbf{z}_p^{(n)}\} \quad (3.22)$$

La projection tenseur en vecteur formée avec ces P projections multilinéaires élémentaires $\{\mathbf{z}_p^{(n)}\}_N^P$ projette \mathbf{X}_m sur un vecteur $\mathbf{y}_m \in \mathbb{R}^{P \times 1}$ comme

$$\mathbf{y}_m = \mathbf{X}_m \times_{n=1}^N \{\mathbf{z}_p^{(n)}\} \quad (3.23)$$

TROD vise à minimiser l'erreur $\varepsilon_{\text{TROD}}$ de reconstruction par le moindres carrés entre les tenseurs d'origine et la reconstruction (ou approximation) à partir d'une combinaison linéaire d'un ensemble de tenseurs de rang un comme suit

$$\varepsilon_{\text{TROD}} = \sum_{m=1}^M \left\| \mathbf{X}_m - \sum_{p=1}^P (y_{m_p} \cdot \mathbf{z}_p^{(1)} \circ \mathbf{z}_p^{(2)} \circ \dots \circ \mathbf{z}_p^{(N)}) \right\|_F^2 \quad (3.24)$$

Où chaque tenseur de rang un est le produit extérieur des vecteurs de projection d'un Projection multilinéaire élémentaire.

L'évaluation de cette technique sur les bases de données de visage en termes de régression, c'est-à-dire l'efficacité de la technique pour l'approximation fonctionnelle et la classification, a montré que pour la régression, le codage par rang de tenseurs, en tant que technique de réduction de dimensionnalité, dépasse de manière significative les techniques comme la PCA. En ce qui concerne la classification, le codage du tenseur est le meilleur lorsque le nombre des échantillons d'apprentissage est très petit.

3.3.3 Approches d'apprentissage sous-espaces multilinéaires supervisés

Les techniques d'apprentissage supervisé prennent des données étiquetées pour l'apprentissage et génèrent (prédit) l'étiquette des données de test. L'étiquette indique généralement la sortie souhaitée, telle que les étiquettes de classe / groupe pour les données (discrètes). La figure 3.9 donne un aperçu sur les algorithmes d'apprentissage supervisés de sous-espaces multilinéaires.

Nous pouvons distinguer deux types de techniques parmi les approches multilinéaire supervisés: les techniques du projection tenseur en tenseur et les techniques du projection tenseur en vecteur

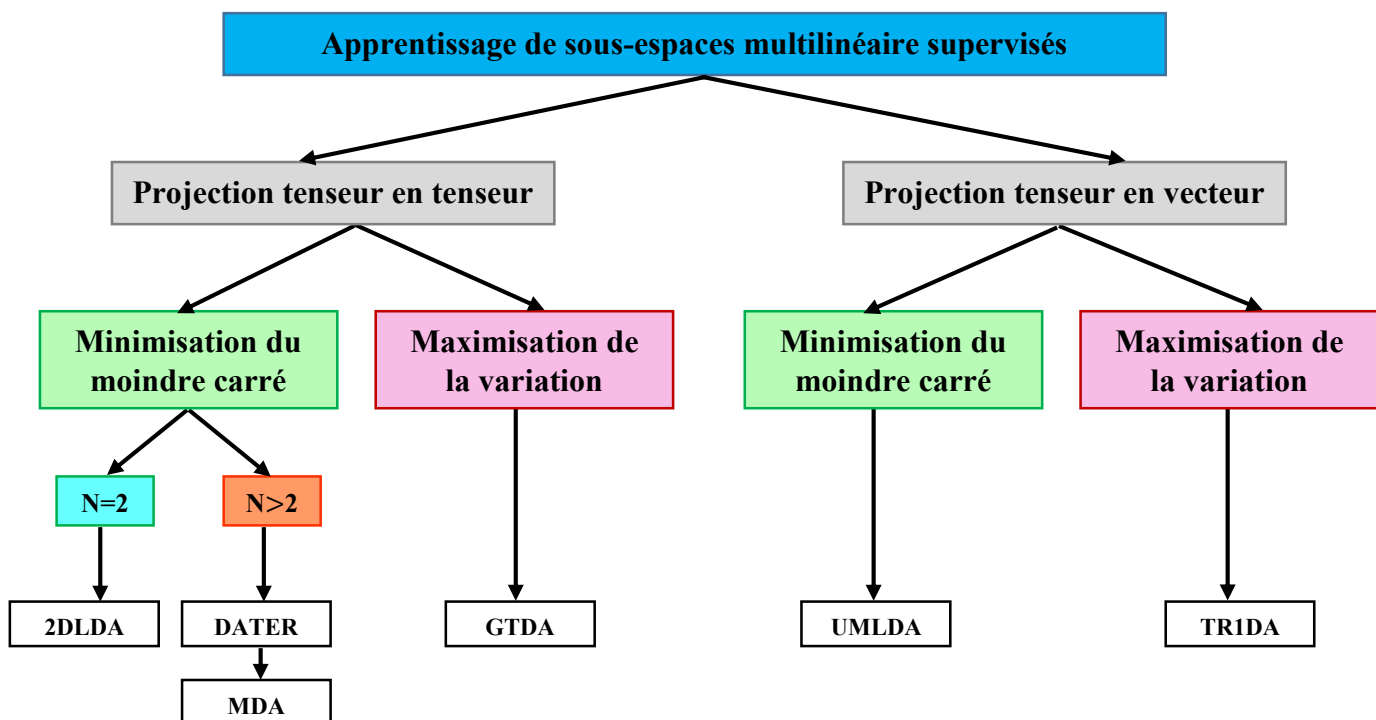


Figure 3.9: Taxonomie des algorithmes d'apprentissage de sous-espaces multilinéaires supervisés.

3.3.3.1 Techniques d'apprentissage de sous-espaces multilinéaires supervisés par la projection tenseur en tenseur

- **2DLDA (Two-Dimensional linear discriminant Analysis)**

Inspiré par 2DPCA, Ye et al [94] ont introduit un algorithme supervisé appelé 2DLDA (Two-Dimensional Linear Discriminant Analysis). Cette méthode vise à maximiser la dispersion inter-classe tout en minimisant la dispersion intra-classe dans le sous-espace projeté comme dans LDA classique. 2DLDA conserve la structure matricielle originale des données. Cependant, 2DLDA ne prend pas en compte la structure multiple des données. Cet algorithme résout deux transformations linéaires (matrices de projection) $Y^{(1)} \in \mathbb{R}^{I_1 \times P_1}$ ($P_1 < I_1$) et $Y^{(2)} \in \mathbb{R}^{I_2 \times P_2}$ ($P_2 < I_2$) qui projettent une image $A_k \in \mathbb{R}^{P_1 \times I_2}$ à une image de faible dimension V_k

$$V_k = Y^{(1)T} A_k Y^{(2)} = A_k \times_1 Y^{(1)T} \times_2 Y^{(2)T} \in \mathbb{R}^{P_1 \times P_2} \quad (3.25)$$

De telle sorte que le rapport entre la dispersion inter-classe S_B et la diffusion intra-classe S_W est maximisé, où S_B et S_W sont définis comme

$$S_B = \sum_{c=1}^D N_c \|\bar{A}_c - \bar{A}\|_F^2 = tr(\sum_{c=1}^D M_c (\bar{A}_c - \bar{A})(\bar{A}_c - \bar{A})^T) \quad (3.26)$$

$$S_W = \sum_{c=1}^D N_c \|\bar{A}_{c_k} - \bar{A}_{c_k}\|_F^2 = tr(\sum_{k=1}^M (\bar{A}_k - \bar{A}_{c_k})(\bar{A}_k - \bar{A}_{c_k})^T) \quad (3.27)$$

Respectivement. Dans ces définitions N_c est le nombre de classes, M_c est le nombre d'échantillons pour chaque classe c , c_k étiquette de classe pour le $m^{ième}$ échantillon d'objet, \bar{A}_c est l'image moyen pour chaque classe, \bar{A} est l'image moyenne pour toutes les données d'apprentissage. La fonction objective de 2DLDA écrit comme suit

$$\{U^{(1)}, U^{(2)}\} = \arg \max_{U^{(1)}, U^{(2)}} \frac{\sum_{c=1}^D M_c \|\bar{V}_c - \bar{V}\|_F^2}{\sum_{k=1}^M \|V_k - \bar{V}_{c_m}\|_F^2} \quad (3.28)$$

Le 2DLDA est testé et évalué dans [94] sur la base de données de visage AT&T. Les résultats expérimentaux montrent que le 2DLDA est plus efficace et plus performant en calcul que les algorithmes actuels du LDA pour l'extraction et la reconnaissance des caractéristiques faciales.

- **DATER (Discriminant Analysis with Tensor Representation)**

Plus tard, DATER (Discriminant Analysis with Tensor Representation) a été proposé par Yan et al [108] comme une extension multilinéaire de la méthode LDA et d'ordre supérieur de 2DLDA. Ceci vise à mettre en œuvre l'analyse discriminante directement sur les données tensorielles naturelles en maximisant le taux de dispersion. Comme avec MPCA, l'algorithme DATER ne peut pas résoudre le problème d'optimisation de manière déterministe pour la projection tenseur en tenseur. Ainsi, une procédure de projection itérative est suivie de la même manière que dans MPCA et dans toute autre approche de sous-espace multilinéaire typique.

M matrices de projection $Y^{(k)} \in \mathbb{R}^{I_k \times P_k}$ sont obtenus pour tous $k=1,2,\dots, M$, qui a utilisé plus tard pour projeter chaque tenseur d'entrée d'ordre M $\mathbf{A}_m \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M}$ dans un tenseur réduit $\mathbf{X}_m \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M}$, où $P_k < I_k$ pour tous $k=1,2,\dots, M$. Cette méthode formule le critère objectif de discrimination du tenseur et maximise le rapport entre la dispersion inter-classe S_B et la dispersion intra-classe S_W .

$$S_B = \sum_{c=1}^D M_c \|\bar{\mathbf{X}}_c - \bar{\mathbf{X}}\|_F^2 \quad (3.29)$$

$$S_W = \sum_{m=1}^M \|\mathbf{X}_m - \bar{\mathbf{X}}_{C_m}\|_F^2 \quad (3.30)$$

Où D est le nombre de classes, M_c est le nombre d'échantillons pour chaque classe c, c_m étiquette de classe pour le $m^{\text{ième}}$ échantillon d'objet, \mathbf{X}_m est l'échantillon caractéristique extrait du tenseur d'entrée \mathbf{A}_m , $\bar{\mathbf{X}}$ est le tenseur moyen pour toutes les données d'apprentissage, et $\bar{\mathbf{X}}_c$ est le tenseur moyen pour chaque classe c. La fonction objective de DATER écrit comme suit

$$U_K^* = \arg \max_{U_K} \frac{Tr(U_K^T S_B U_K)}{Tr(U_K^T S_W U_K)} \quad (3.31)$$

La transformation est ensuite déterminée en dérivant les vecteurs propres résultant de la dispersion entre classes et de la dispersion intra-classe. La figure 3.10 présente l'évolution du critère objectif de la méthode DATER au fil des itérations.

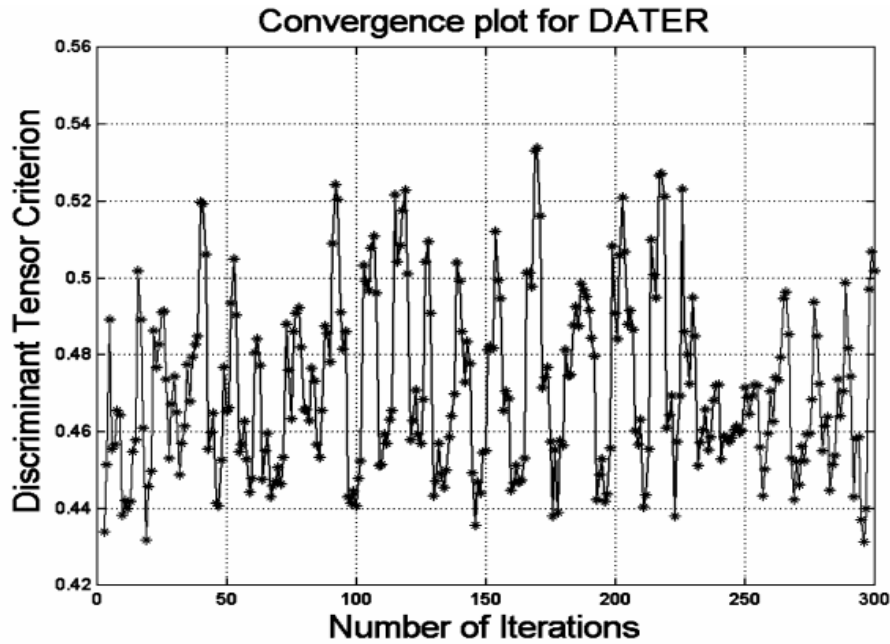


Figure 3.10: L'évolution du critère objectif de la méthode DATER au fil des itérations [108].

La méthode DATER a été appliquée avec succès dans les applications de reconnaissance faciale [96, 108]. Similaire à la MPCA, le DATER conserve uniquement la structure globale des échantillons de données tensorielles au prix de la perte d'informations significatives concernant la géométrie locale des voisins.

- **GTDA (General Tensor Discriminant Analysis)**

Similaire à DATER, le GTDA (General Tensor Discriminant Analysis) qui a été proposé par Tao et al [95] résout pour les transformations multilinéaires $\{Y^{(k)} \in \mathbb{R}^{I_k \times P_k}, P_k < I_k, K = 1, \dots, N\}$ qui projette un tenseur $\mathbf{A}_m \in \mathbb{R}^{I_1 \times \dots \times I_N}$ à un tenseur de faible dimension $\mathbf{X}_m \in \mathbb{R}^{P_1 \times \dots \times P_N}$. La différence avec DATER est qu'elle maximise le critère de différence de dispersion (plutôt que le ratio de dispersion) et que les matrices de projection dans GTDA ont des colonnes orthonormées. GTDA vise à maximiser une extension multilinéaire du critère discriminant basé sur la différence de dispersion dans [109]. Contrairement à 2DLDA et DATER, cet algorithme a une bonne propriété de convergence et le premier algorithme discriminant qui converge vers une solution locale. La fonction objective de GTDA est définie comme suit :

$$\begin{aligned}
 \{Y^{(n)}\} &= \arg \max_{Y^{(n)}} S_B - \xi \cdot S_W \\
 &= \arg \max_{Y^{(n)}} \sum_{c=1}^D M_c \|\mathbf{Y}_c - \bar{\mathbf{Y}}\|_F^2 - \xi \sum_{m=1}^M \|\mathbf{Y}_m - \bar{\mathbf{Y}}_{c_m}\|_F^2
 \end{aligned} \tag{3.32}$$

Où ξ est un paramètre de réglage.

- **MDA (Multilinear Discriminant Analysis)**

Plus tard, Yan et al ont présenté une nouvelle approche nommée MDA (Multilinear Discriminant Analysis) [21] qui est connu comme l'extension du tenseur de la méthode classique LDA. Cet algorithme résout le problème de réduction de dimensionnalité supervisée en codant un objet image sous la forme d'un tenseur d'ordres deux ou même supérieur. En général, MLDA recherche une projection multilinéaire qui mappe les données d'entrée d'un espace à un autre espace de dimension inférieure et plus discriminant. La fonction objective de MDA est

$$(U_1^*) = \arg \max_{U^k} \frac{\sum_{i=1}^C P_i \|\bar{A}_i \times_k Y_k - \bar{A} \times_k Y_k\|^2}{\|A_j \times_k Y_k - \bar{A}_{c_j} \times_k Y_k\|^2} \quad (3.33)$$

Où \bar{A}_i est le tenseur moyen des échantillons appartenant à la classe i , \bar{A} est le tenseur moyen total de tous les échantillons, et P_i est le nombre d'échantillon de la classe i .

Parce qu'il n'y a pas de solution sous forme fermée pour (3.32), un algorithme itératif permettant de trouver les projections optimales locales a donc été proposé dans [21]. Dans chaque itération, $Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_n$ sont supposées connues et les échantillons d'image sont projetés sur ces matrices de projection. Alors le problème d'optimisation peut être reformulé en un problème d'analyse discriminante spécial comme :

$$Y_k^* = \arg \max_{U^k} \frac{\text{tr}(Y_k^T S_B^k Y_k)}{\text{tr}(Y_k^T S_W^k Y_k)} \quad (3.34)$$

Où S_B^k et S_W^k sont les matrices de dispersion inter-classes et intra-classes en mode K définies comme suit :

$$S_B^k = \sum_{i=1}^C \sum_{j=1}^{\Pi_{o \neq k} m_o} p_c (\bar{a}_c^k(j) - \bar{a}^k(j)) (\bar{a}_c^k(j) - \bar{a}^k(j))^T \quad (3.35)$$

$$S_W^k = \sum_{i=1}^N \sum_{j=1}^{\Pi_{o \neq k} m_o} p_i (a_i^k(j) - \bar{a}_{c_i}^k(j)) (a_i^k(j) - \bar{a}_{c_i}^k(j))^T \quad (3.36)$$

Où, pour la facilité de présentation, \mathbf{a}_i^k représente le vecteur de la $j^{\text{ième}}$ colonne de la matrice A_i^k qui est la matrice dépliée en mode k du tenseur de l'échantillon \mathbf{A}_i , $\bar{\mathbf{a}}_c^k(j)$ et $\bar{\mathbf{a}}^k(j)$ sont définis de la même manière que \mathbf{a}_i^k vis-à-vis les tenseurs $\bar{\mathbf{A}}_c$ et $\bar{\mathbf{A}}$. La figure 3.11 donne une illustration sur la variance inter-classes et la variance intra-classes des données de visage avant et après l'utilisation la méthode MDA.

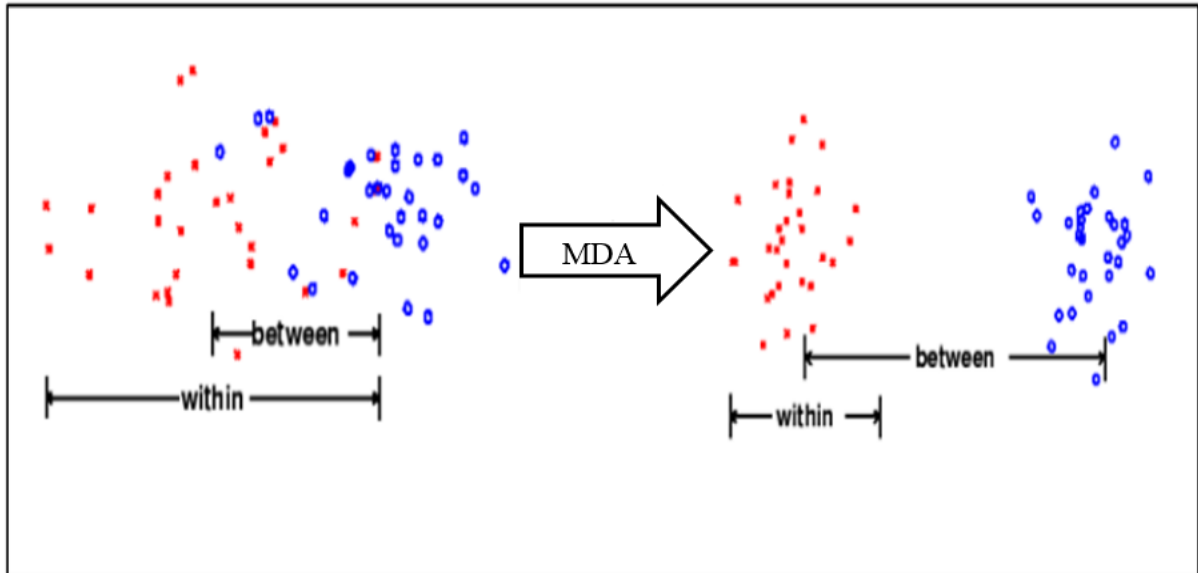


Figure 3.11: Variance inter-classe et variance intra-classe avant et après l'utilisation MDA.

Des expériences approfondies dans [21] sur les bases de données ORL, CMU PIE et FERET en codant des images de visage en tant que tenseurs du deuxième ou troisième ordre démontré ont montré que l'algorithme MDA proposé basé sur des tenseurs supérieurs a le potentiel de surperformer les algorithmes traditionnels d'apprentissage par le sous-espace basés sur des vecteurs tels que PCA et LDA, en particulier dans les cas avec des échantillons de petite taille.

3.3.3.2 Techniques d'apprentissage de sous-espaces multilinéaires supervisées par la projection tenseur en vecteur

- **TR1DA (Tensor Rank-One Discriminant Analysis)**

Tao et al [64] ont introduit une nouvelle approche supervisée pour l'apprentissage discriminant utilisant des représentations tensorielles nommé TR1DA (Tensor Rank-One Discriminant Analysis). Le TR1DA est la première méthode supervisée basée sur la projection tenseur en vecteur qui projette un tenseur sur un vecteur de dimension faible tout en maximisant le critère

de différence de dispersion scalaire. La méthode doit déterminer un paramètre de réglage comme dans GTDA. De plus, cet algorithme s'appuie sur les résidus calculés de manière répétée utilisés dans TROD [110], ce qui convient mieux à l'approximation tenseur que pour la discrimination. TR1DA trouve P projections multilinéaires élémentaires $\mathbf{z}_p^{(n)}$ à la fois comme dans TROD et UMPCA qui projettent un tenseur $\mathbf{X}_m \in \mathbb{R}^{I_1 \times \dots \times I_N}$ en scalaire y_{kp} .

$$\mathbf{y}_{m_p} = \mathbf{X}_m \times_1 \mathbf{z}_p^{(1)T} \times_2 \mathbf{z}_p^{(2)T} \dots \times_N \mathbf{z}_p^{(N)T} = \mathbf{X}_m \times_{n=1}^N \{\mathbf{z}_p^{(n)}\} \quad (3.37)$$

La projection tenseur en vecteur est formée avec ces P projections multilinéaires élémentaires $\{\mathbf{z}_p^{(n)}\}_N^P$ projettent \mathbf{X}_m sur un vecteur $\mathbf{y}_m \in \mathbb{R}^{P \times 1}$ comme :

$$\mathbf{y}_m = \mathbf{X}_m \times_{n=1}^N \{\mathbf{z}_p^{(n)}\} \quad (3.38)$$

La dispersion inter-classe S_B de $\{\mathbf{y}_{m_p}\}$ pour le $p^{\text{ième}}$ projection multilinéaires élémentaires est

$$S_{B_p} = \sum_{c=1}^C M_c (\bar{y}_{c_p} - \bar{y}_p)^2 \quad (3.39)$$

La dispersion intra-classe S_W de $\{\mathbf{y}_{m_p}\}$ pour le $p^{\text{ième}}$ projection multilinéaires élémentaires est

$$S_{W_p} = \sum_{m=1}^M (y_{m_p} - \bar{y}_{C_{k_p}})^2 \quad (3.40)$$

Où C est le nombre de classes, M_c est le nombre d'échantillons pour la classe c, m est l'étiquette de classe pour le $m^{\text{ième}}$ échantillon d'apprentissage, \bar{y}_p est la moyenne globale définie comme $\bar{y}_p = \frac{1}{M} \sum_m y_{m_p} = 0$ et \bar{y}_{c_p} est la moyenne de classe définie comme $\bar{y}_{c_p} = \frac{1}{M_c} \sum_{m, c_m = c} y_{m_p}$

La fonction objective de TR1DA peut être écrite comme

$$\begin{aligned} \mathbf{z}_p^{(n)} &= \arg \max_{\mathbf{z}_p^{(n)}} S_{B_p} - \xi \cdot S_{W_p} \\ &= \arg \max_{\mathbf{z}_p^{(n)}} \sum_{c=1}^C M_c (\bar{y}_{c_p} - \bar{y}_p)^2 - \xi \sum_{m=1}^M (y_{m_p} - \bar{y}_{C_{k_p}})^2 \end{aligned} \quad (3.41)$$

Où ξ est un paramètre de réglage.

Basé sur beaucoup d'expériences dans [111] sur les bases de données de visage PIE et Yale, la méthode TR1DA a démontrée ses avantages de manière significative par rapport les méthodes traditionnelles de sous-espace, tel que PCA, LDA, LPP.

- **UMDLA (Uncorrelated Multilinear Discriminant Analysis)**

L'algorithme UMDLA (Uncorrelated Multilinear Discriminant Analysis) est une extension de la méthode linéaire ULDA (Uncorrelated Linear Discriminant Analysis) [112] formulée par Lu et al [96]. Il extrait les caractéristiques discriminantes non corrélées directement à partir des données de tenseur en résolvant une projection tenseur en vecteur qui maximise un critère basé sur le rapport de dispersion typique. Pour améliorer les performances dans le scénario de petite taille d'échantillon, un facteur de régularisation adaptatif est incorporé afin de réduire la variance de l'estimation de la dispersion intra-classe par le biais d'un paramètre de régularisation indépendant des données, générant ainsi la UMLDA régularisé (R-UMLDA). Le système proposé par Xu et al de reconnaissance typique basée sur la méthode UMLDA est présenté dans la figure 3.12.

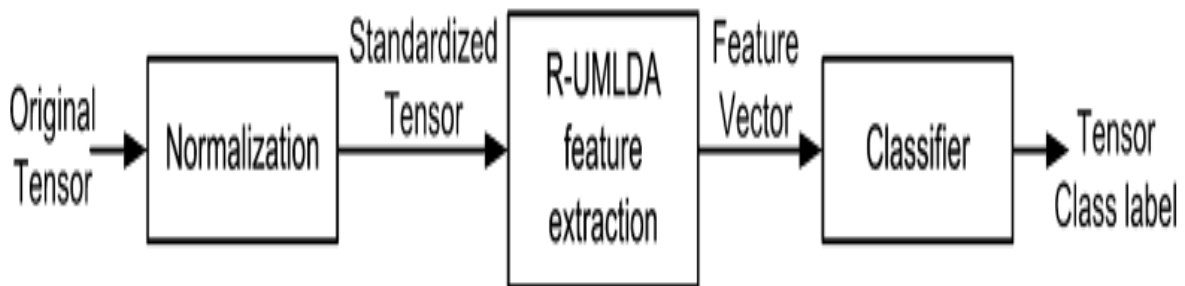


Figure 3.12: Reconnaissance des objets (visages) tenseurs basés sur UMLDA [96].

Comme dans UMPCA et TROD, les $p^{i\text{ème}}$ caractéristiques projetées (scalaires) sont $\{y_{k_p}, k = 1, \dots, M\}$ où M est le nombre d'échantillons d'apprentissage et y_{m_p} est la projection du $m^{i\text{ème}}$ échantillon A_m par la projection multilinéaire élémentaire $\{\mathbf{z}_p^{(n)T}, n = 1, \dots, N\}$. La figure 3.13 illustre une représentation de codage de l'information discriminante trouvée dans une partie du visage avec les premières projections multilinéaires élémentaires par UMLDA

$$\mathbf{y}_{m_p} = A_m \times_{n=1}^N \{\mathbf{z}_p^{(n)T}, n = 1, \dots, N\} \quad (3.42)$$

Leur dispersion inter-classe S_B et la dispersion intra-classe S_W Sont alors

$$S_{B_p} = \sum_{c=1}^C M_c (\bar{y}_{c_p} - \bar{y}_p)^2 \quad (3.43)$$

$$S_{W_p} = \sum_{m=1}^M (y_{m_p} - \bar{y}_{c_{m_p}})^2 \quad (3.44)$$

Où $\bar{y}_p = \frac{1}{M} \sum_m y_{m_p} = 0$ et $\bar{y}_{c_p} = \frac{1}{M_c} \sum_{m, c_{m_p} = c} y_{m_p}$. Ainsi, le critère de discrimination de fisher pour les échantillons de scalaire $p^{\text{ième}}$ est

$$F_p = \frac{S_{B_p}}{S_{W_p}} \quad (3.45)$$

La fonction objective de la $p^{\text{ième}}$ projection multilinéaire élémentaire définit comme suit

$$\{z_p^{(n)T}, n = 1, \dots, N\} = \operatorname{argmax} F_p \quad (3.46)$$

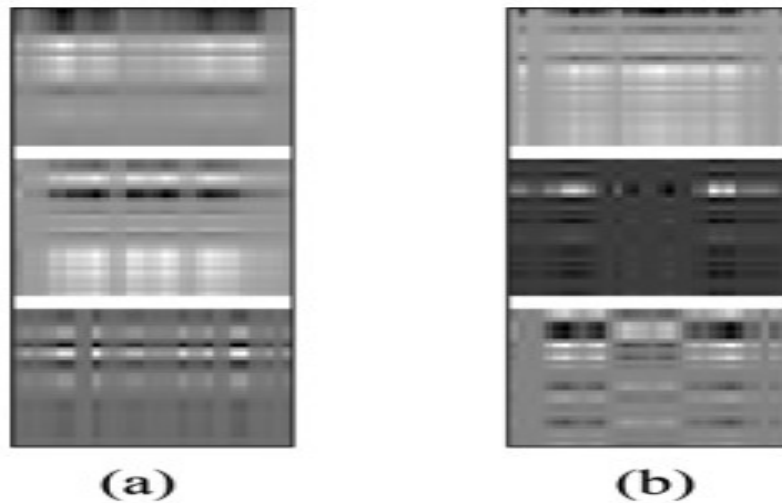


Figure 3.13: Illustration des trois premières projections multilinéaires élémentaires par UMLDA de (a) la base de données PIE, (b) la base de données FERET [96].

Des expériences sur des applications biométriques faciales, testées sur les bases de données de visage PIE et FERET dans [96], démontrent que le dispositif de reconnaissance basé sur UMLDA obtient les meilleurs résultats globaux par rapport aux systèmes de reconnaissance

basés sur d'autres solutions de sous-espaces multilinéaires (MPCA, DATER, GTDA, TR1DA) ainsi que des algorithmes de sous-espaces linéaires (Bayesian [113], LDA, ULDA).

3.4 Conclusion

Dans ce chapitre, nous avons minutieusement rappelé les méthodes d'apprentissage de sous-espaces multilinéaire pour l'analyse des données multidimensionnelles. Dans un premier temps, nous avons commencé par l'historique d'apprentissage multilinéaire. Ensuite, nous avons présenté les techniques existantes de sous-espaces multilinéaires dans la littérature pour la reconnaissance du visage, selon deux catégories principales : les approches supervisées et les approches non supervisées. Les approches d'apprentissage de sous-espaces multilinéaire permettent de traiter une quantité énorme de données et de réduire la dimensionnalité des données multidimensionnelles directement à partir de leurs représentations tensorielles naturelles plutôt que les converties en vecteurs. Elles fournissent également un moyen efficace d'extraire des caractéristiques utiles qui sont meilleures que celles des méthodes d'apprentissage de sous-espaces linéaire. De plus, elles préservent la structure naturelle des données multidimensionnelles. En conséquence, plusieurs chercheurs ont proposé de meilleures formulations mathématiques pour améliorer la performance des algorithmes de sous-espaces multilinéaire. Nous pouvons également continuer ce travail sur le développement des méthodes basées sur l'analyse multidimensionnelle afin que la robustesse des systèmes de reconnaissance faciale puisse être augmentée.

Chapitre 4

MEFDA pour la vérification multimodale du visage 2D+3D robuste

4.1 Introduction	70
4.2 MEFDA (Multilinear Enhanced Fisher Discriminant Analysis).....	73
4.3 Aperçu de l'approche de vérification de visage proposé	78
4.3.1 Conception du tenseur basé sur les caractéristiques locales.....	78
4.3.1.1 Descripteur LPQ (Local Phase Quantization).....	79
4.3.2 Réduction de la dimensionnalité et classification.....	82
4.4 Evaluation expérimentale.....	83
4.4.1 Base de données de visage.....	84
4.4.1.1 Base de données FRGC V2.0.....	84
4.4.1.2 Base de données CASIA 3D	84
4.4.1.3 Base de données Bosphorus.....	85
4.4.2 Paramètres de réglage.....	85
4.4.3 Résultats et discussion	86
4.4.3.1 Vérification uni-modale de visage 2D et 3D	86
4.4.3.2 Vérification multimodal de visage 2D + 3D.....	87
4.4.3.3 MEFDA vis à vis EFM	88
4.4.3.4 Le temps de complexité	90
4.4.3.5 MEFDA vis à vis toutes les approches	90
4.4.3.6 Comparaison avec l'état de l'art.....	92
4.5 Conclusion.....	94

4.1 Introduction

Au cours de la dernière décennie, la reconnaissance automatique des visages est devenue un terrain fertile pour les chercheurs en vision par ordinateur et en reconnaissance de formes, en raison de sa grande importance et de ses diverses applications. La représentation des données faciales, l'extraction de caractéristiques, la projection de sous-espaces et la classification sont les étapes principales de tout système de reconnaissance faciale. Plusieurs approches ont été proposées à chaque étape pour améliorer les performances de reconnaissance contre des défis tels que des éclairages, des poses, des expressions, des occlusions... etc. [114-118]. De plus, l'utilisation de tenseurs nécessite le passage des techniques linéaires aux techniques multilinéaires. Dans ce qui suit, nous présentons brièvement les travaux de littérature les plus importants sur la réduction de dimensionnalité multilinéaire pour l'analyse de tenseurs et leurs applications à la reconnaissance faciale, ainsi que sur l'utilisation de données de visage 2D et 3D pour augmenter la précision de reconnaissance faciale.

Le processus de réduction de la dimensionnalité est l'une des tâches les plus importantes dans plusieurs domaines de recherche, tels que la reconnaissance de formes, l'apprentissage automatique, l'exploration de données et l'analyse de visage [107, 119, 120]. En effet, la réduction de dimensionnalité permet d'analyser les données plus facilement et plus rapidement pour l'apprentissage automatique en transformant un espace de grande dimension en un sous-espace de plus petite dimension tout en conservant des données précieuses et en supprimant les données redondantes ou inutiles. Les méthodes classiques de réduction de dimensionnalité linéaire assurent la projection des données dans un autre sous-espace dans lequel la dimension d'origine de ces données est réduite. Parmi les plus célèbres de ces méthodes on trouve, PCA [121], LDA [71], Whitened PCA [122], EDA (Exponential Discriminant Analysis) [123]. Ces méthodes linéaires et non linéaires ne peuvent manipuler qu'avec la représentation de données vectorisée, tandis que les données de nombreuses applications telles que la reconnaissance faciale dans la nature sont plus naturellement représentés comme des objets multilinéaires basés sur des tenseurs d'ordre élevé [96, 107, 117, 118, 120, 124]. Plusieurs défis de la vérification du visage dans la nature, tels que des éclairages, des expressions et des poses peuvent être modélisés comme un problème d'algèbre multilinéaire dans lequel les caractéristiques, les personnes et les différents défis sont considérés comme des modes de données tensorielles. Cette représentation tenseur des données faciales a conduit les chercheurs à proposer plusieurs méthodes de réduction de dimensionnalité multilinéaires telles que Multilinear PCA [24], MDA (Multilinear Discriminant Analyses) [125], Multilinear Whitened PCA [117], TEDA

(Tensor Exponential Discriminant Analysis) [117]. Les méthodes de sous-espace linéaire établies ont été généralisées au cas multilinéaire. Par exemple, la MPCA a généralisé la PCA traditionnelle aux espaces tenseurs. MPCA définit une projection multilinéaire des données de tenseur d'origine dans un sous-espace de tenseur de dimension inférieure en conservant les directions de variation maximales dans les données d'origine. Cependant, comme PCA, l'inconvénient principal de la MPCA est sa nature non supervisée, où la projection des données est moins efficace en ce qui concerne le défi de la séparation des classes. Pour résoudre ce problème, LDA peut être appliqué directement après MPCA. Yan et al ont étendu la LDA à la MDA pour traiter des données de représentations de tenseurs d'ordre supérieur afin de donner un pouvoir plus discriminant. Pour réduire les dimensions des tenseurs d'ordre élevé, MDA applique un processus itératif appelé optimisation en mode k afin de maximiser un critère discriminant basé sur le tenseur également pour l'apprentissage itératif de nombreux sous-espaces discriminants interdépendants. Lu et al ont proposé l'algorithme Uncorrelated MDA [96], qui est basé sur l'extraction des caractéristiques non corrélées en résolvant la projection tenseur en vecteur. Plus récemment, des résultats encourageants en reconnaissance de visage 2D et 3D sont obtenus par Ouamane et al [117] dans lequel, EDA et WPCA sont étendus à TEDA et WMPCA, respectivement, ce qui nous rend plus motivés pour travailler sur ce problème.

Par ailleurs, de nombreux algorithmes ont été développés pour la combinaison des données de visages 2D et 3D afin d'augmenter la précision de la vérification de visage en milieux non contrôlés. Xu et al [126] ont proposé un modèle de reconnaissance des visages basé sur les réponses en ondelettes de Gabor extraites à la fois d'images 3D de profondeur et d'images 2D d'intensité. Les caractéristiques de Gabor de chaque modalité sont projetées dans un sous-espace inférieur à l'aide d'une combinaison de LDA et l'Adaboost. Al-Osaimi et al [127] ont proposé une fusion optimisée spatialement au niveau des données pour la reconnaissance des visages 2D et 3D. Plusieurs fonctions de fusion non linéaire sont optimisées en tenant compte des expressions du visage et des variations d'éclairage. Lu et al [18] ont proposé l'algorithme LMKMML (Localized Multi-Kernel Multi-Metric Learning) pour la fusion de différentes caractéristiques statistiques du visage. LMKMML apprend conjointement plusieurs fonctionnalités et métriques de distance dans les espaces du noyau, favorisant ainsi des données complémentaires de la reconnaissance faciale.

Comme mentionné ci-dessus, LDA est considérée comme l'une des méthodes de projection linéaires les plus utilisées dans la reconnaissance faciale en raison de son pouvoir discriminant

accordé aux différentes classes de la base de données [118, 126, 128]. L'algorithme EFM [129] est une extension de l'algorithme LDA qui augmente la capacité de généralisation de LDA basée sur la diagonalisation simultanée des deux matrices de dispersion (les matrices intra-classes et inter-classes). Dans ce chapitre, nous abordons le problème de la vérification de visage en 2D+3D dans la nature en basant sur la représentation tensorielle des images de visage.

En particulier, l'originalité de ce travail peut être expliquée dans les points suivants :

- Proposition d'une nouvelle méthode de réduction de dimensionnalité multilinéaire supervisée appelée MEFDA (Multilinear Enhanced Fisher Discriminant Analysis), pour la réduction de la dimensionnalité et la classification des données de tenseurs d'ordre élevé, dans lesquelles la marge entre échantillons appartenant à des classes différentes est agrandie, tandis que la marge entre échantillons appartenant aux mêmes classes est réduite. De plus, MEFDA cherche à prendre en compte ces deux avantages, i) le bon pouvoir discriminant de EFM et ii) le côté compétitif des données de représentation tensorielles, dans lesquelles la structure d'origine et la géométrie naturelle des données faciales sont respectées.
- Proposition d'une nouvelle représentation basée sur les tenseurs d'ordre élevé. Cette représentation combine différentes échelles descriptives locales permettant une meilleure discrimination.
- Conception d'une nouvelle approche de vérification automatique de visages 2D + 3D en utilisant des caractéristiques locales 2D et 3D basées sur la représentation des tenseurs d'ordre élevé.
- Evaluation de manière empirique l'approche proposée pour la vérification multimodale de visage sur trois bases de données de visage difficiles (FRGC v2.0, CASIA 3D et Bosphorus). La comparaison avec les méthodes de l'état de l'art démontre l'efficacité de notre approche

Le reste de ce chapitre est prévu comme suit. Dans la section 2, la méthode proposée MEFDA (Multilinear Enhanced Fisher Discriminant Analysis) est formulée. Dans la section 3, l'aperçu de notre système de vérification des visages est donné. Dans la section 4, les résultats expérimentaux réalisés sur les trois bases de données de visage sont décrits. Enfin, nous terminerons ce chapitre par une conclusion et des travaux futurs.

4.2 MEFDA (Multilinear Enhanced Fisher Discriminant Analysis)

Dans cette section, la méthode EFM (Enhanced Fisher Model) est d'abord expliquée brièvement. Par la suite, la MEFDA proposée est décrite avec des formulations mathématiques détaillées.

EFM est développé à partir de la LDA traditionnelle avec une meilleure capacité de généralisation dans la séparation discriminante des classes ciblées, par la diagonalisation des deux matrices de dispersion inter-classes et intra-classes. De plus, cela préserve un équilibre approprié entre le choix des valeurs propres et la nécessité que les valeurs propres de la matrice de dispersion intra-classe ne soient pas trop petites.

Soit $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$ l'ensemble des échantillons d'apprentissage appartenant aux classes C , les principales étapes de l'algorithme EFM peuvent être résumées comme suit:

- Etape 1: Trouver les matrices de dispersion intra-classe S_w et inter-classe S_b

$$S_w = \sum_{i=1}^C \frac{1}{c_i} \sum_{j=1}^{c_i} (\mathbf{a}_j^i - \mathbf{d}^i) (\mathbf{a}_j^i - \mathbf{d}^i)^T \quad (4.1)$$

$$S_b = \sum_{i=1}^M (\mathbf{d}^i - \bar{\mathbf{d}}) (\mathbf{d}^i - \bar{\mathbf{d}})^T \quad (4.2)$$

où c_i est le nombre d'échantillons de classe i , \mathbf{d}^i est la moyenne de la classe i , et $\bar{\mathbf{d}}$ est la moyenne de tous les échantillons de données.

- Etape 2: Calculer les valeurs propres X et les vecteurs propres V de la matrice intra-classe S_w .
- Etape 3: Calculer la nouvelle matrice inter-classe N_b

$$N_b = X^{-1/2} X^T S_b V X^{-1/2} \quad (4.3)$$

- Etape 4: Calculer les valeurs propres W et les vecteurs propres U de la matrice de N_b .
- Etape 5: Calculer la matrice de projection globale T_p

$$T_p = V X^{-1/2} U \quad (4.4)$$

Dans le cadre EFM, nous étendons cet algorithme pour manipuler les données de tenseur d'ordre élevé. Le tenseur d'apprentissage $\mathbf{Z} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_L}$ appartenant à C différentes classes et

chaque classe \mathbf{Z}_j contient des échantillons. Ainsi, l'ensemble de données d'apprentissage de la galerie est représenté comme un tenseur d'ordre $(L+1)$ $\tilde{\mathbf{Z}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_L \times M}$, où M est le nombre d'échantillons d'apprentissage.

Pour trouver une transformation multilinéaire, le critère du tenseur discriminant [125] est conçu pour rechercher plusieurs sous-espaces de projection interdépendants, qui minimisent la dispersion intra-classe et maximisent la dispersion inter-classe en même temps, Le critère du tenseur discriminant est présenté de la manière suivant:

$$U_{k|k=1}^* = \arg \max_{U_{k|k=1}^L} \frac{\sum_{j=1}^C C_j \|\bar{\mathbf{Z}}_j \times_1 U_1 \cdots \times_L U_L - \bar{\mathbf{Z}} \times_1 U_1 \cdots \times_L U_L\|^2}{\sum_{i=1}^M \|\mathbf{z}_i \times_1 U_1 \cdots \times_L U_L - \bar{\mathbf{Z}}_{c_i} \times_1 U_1 \cdots \times_L U_L\|^2} \quad (4.5)$$

Où : $\bar{\mathbf{Z}}_j$ est le tenseur moyen pour chaque classe j , $\bar{\mathbf{Z}}$ est le tenseur moyen pour toutes les données d'apprentissage. L'équation 4.5, représente un problème d'optimisation avec une contrainte non linéaire d'ordre élevé (il n'existe pas de solution de forme fermée). Heureusement, nous pouvons utiliser l'approche d'optimisation itérative comme dans [125] pour obtenir les sous-espaces discriminants interdépendants. A chaque itération, L matrices de projection $U_1, \dots, U_{k-1}, U_{k+1}, \dots, U_L$ sont supposées connues et la matrice de projection U_k est calculée par le critère d'optimisation du tenseur en mode k suivant:

$$U_k^* = \arg \max_{U_k} \frac{\sum_{j=1}^C C_j \|\bar{\mathbf{Z}}_j \times_k U_k - \bar{\mathbf{Z}} \times_k U_k\|^2}{\sum_{i=1}^M \|\mathbf{z}_i \times_k U_k - \bar{\mathbf{Z}}_{c_i} \times_k U_k\|^2} \quad (4.6)$$

Le problème d'optimisation dans (4.6) est résolu par une solution naturelle, le tenseur est déplié en matrices dans le mode k par l'équation (2.1). Le problème d'optimisation pourrait être transformé par une analyse discriminante exceptionnelle comme suit:

$$U_k^* = \arg \max_{U_k} \frac{\text{Tr}(U_k^T S_b^k U_k)}{\text{Tr}(U_k^T S_w^k U_k)} \quad (4.7)$$

où S_b^k et S_w^k sont les matrices de dispersion inter-classe et intra-classe pour le mode k , respectivement.

$$S_b^k = \sum_{y=1}^{\prod_{i \neq k} I_i} S_b^y, S_b^y = \sum_{j=1}^C C_j (\bar{\mathbf{z}}_j^{k,y} - \bar{\mathbf{z}}^{k,y}) (\bar{\mathbf{z}}_j^{k,y} - \bar{\mathbf{z}}^{k,y})^T \quad (4.8)$$

$$S_w^k = \sum_{y=1}^{\prod_{i \neq k} I_i} S_w^y, S_w^y = \sum_{i=1}^M \left(\mathbf{z}_i^{k,y} - \bar{\mathbf{z}}_c^{k,y} \right) \left(\mathbf{z}_i^{k,y} - \bar{\mathbf{z}}_c^{k,y} \right)^T \quad (4.9)$$

où, pour toutes les présentations, $\mathbf{z}_i^{k,y}$ est le $y^{\text{ième}}$ vecteur colonne de la matrice dépliée en k -mode Z_i^k du tenseur d'échantillon \mathbf{Z}_i , $\bar{\mathbf{z}}_j^{k,y}$ le $y^{\text{ième}}$ vecteur colonne de la matrice moyenne dépliée en mode k \bar{Z}_j^k de tenseur moyen $\bar{\mathbf{Z}}$ pour chaque classe j , $\bar{\mathbf{z}}^{k,y}$ est représenté de la même manière que $\mathbf{z}_i^{k,y}$ par rapport au tenseur moyen $\bar{\mathbf{Z}}$. Pour chaque mode K , calculer les vecteurs propres et les valeurs propres de la matrice de dispersion intra-classe S_w^k

$$S_w^k = (V_w^k)^T \Lambda_w^k V_w^k \quad (4.10)$$

où V_w^k représente les vecteurs propres et Λ_w^k sont les matrices de valeurs propres.

Maintenant, nous calculons la nouvelle matrice de dispersion inter-classe N_b^k comme suit:

$$N_b^k = (V_w^k)^{-1/2} (\Lambda_w^k)^T S_b^k \Lambda_w^k (V_w^k)^{-1/2} \quad (4.11)$$

Pour chaque mode K , calculer les vecteurs propres et les valeurs propres de la matrice de dispersion inter-classe N_b^k

$$N_b^k = (E_b^k)^T O_b^k E_b^k \quad (4.12)$$

En utilisant la décomposition de vecteurs propres de la matrice N_b^k , le problème d'optimisation de l'équation (4.7) pourrait être remplacé par une solution fermée simple avec analyse discriminante d'ordre supérieur, comme suit:

$$Y^k = \Lambda_w^k V_w^k E^k \quad (4.13)$$

A chaque itération, $Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_L$ sont supposées connues, l'équation (4.5) est convertie en:

$$Y_k^* = \arg \max_{Y_k} \frac{\sum_{j=1}^C C_j \|\bar{\mathbf{z}}_j \times_1 Y_1 \cdots \times_L Y_L - \bar{\mathbf{z}} \times_1 Y_1 \cdots \times_L Y_L\|^2}{\sum_{i=1}^L \|\bar{\mathbf{z}}_i \times_1 Y_1 \cdots \times_L Y_L - \bar{\mathbf{z}}_{C_i} \times_1 Y_1 \cdots \times_L Y_L\|^2} \quad (4.14)$$

Réglage:

$$\mathbf{H}_i = \mathbf{Z}_i \times_i Y_1 \cdots \times_{k-1} Y_{k-1} \times_{k+1} Y_{k+1} \cdots \times_L Y_L \quad (4.15)$$

Alors, l'équation (4.6) devient:

$$Y_k^* = \arg \max_{Y_k} \frac{\sum_{j=1}^C c_j \|\bar{H}_j \times_k Y_k - \bar{H} \times_k Y_k\|^2}{\sum_{i=1}^L \|\mathbf{H}_i \times_k Y_k - \bar{H}_{n_i} \times_k Y_k\|^2} \quad (4.16)$$

Les matrices de projection peuvent être optimisées de manière itérative pour chaque mode k . Le processus itératif de MEFDA se décompose lors de la réalisation de l'une des conditions suivantes: soit i) le nombre d'itérations atteint un maximum prédéfini; ou ii) la différence de la projection estimée entre deux itérations successives est inférieure à un seuil, $\|Y_k^{\text{iteration}} - Y_k^{\text{iteration}-1}\| < I_k \times I_k \times \varepsilon$ où $I_k \times I_k$ est la dimension en mode k et $Y_k^{\text{iteration}}$ est la matrice de vecteurs propres en mode k calculée dans l'équation (4.14)

Les dimensions inférieures finale $I'_1 \times I'_2 \dots \times I'_L$ sont définies en fonction du pourcentage d'énergie (Energy_k) d'élément diagonal de valeurs propres λ_i^k de la matrice des valeurs propres Λ_w^k à conserver pour chaque mode k .

$$\text{Energy}_k = \frac{\sum_{i=1}^{I'_k} \lambda_i^k}{\sum_{i=1}^{I_k} \lambda_i^k} \times 100 \quad (4.17)$$

avec $\lambda_1^k > \lambda_2^k > \dots > \lambda_{I_k}^k$.

La procédure complète pour notre méthode, Multilinear Enhanced Fisher Discriminant Analysis (MEFDA) est fourni dans l'algorithme 1.

Algorithme 1: MEFDA (Multilinear Enhanced Fisher Discriminant Analysis)

Entrées:

- Le tenseur $\tilde{\mathbf{Z}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_L \times M}$ des M échantillons d'apprentissage comprend C classes et chaque classe \mathbf{Z}_j comprend C_j échantillons.
- Le nombre maximal d'itérations: Iteration_{\max}
- L'énergie ($\text{Energie}_1, \text{Energie}_2, \dots, \text{Energie}_L$) utilisée pour choisir les dimensions inférieures finales: $I'_1 \times I'_2 \dots I'_L$

Sorties:

- Les matrices de projections $Y_k = Y_k^t \in \mathbb{R}^{I_k \times I'_k}$, $k = 1, \dots, L$

Algorithme:

1. **Initialisation:** $Y_1^0 = I_{I_1}, Y_2^0 = I_{I_2}, \dots, Y_L^0 = I_{I_L}$
2. **Pour** itr: 1 to Iteration_{\max}
 - a) **Pour** $k=1$ to L
 - ♦ $\mathbf{H}_i = \mathbf{Z}_i \times_1 Y_1^{\text{itr}-1} \dots \times_{k-1} Y_{k-1}^{\text{itr}-1} \times_{k+1} Y_{k+1}^{\text{itr}-1} \dots \times_L Y_L^{\text{itr}-1}$
 - ♦ $H_i^k \leftarrow \mathbf{H}_i$
 - ♦ Calculer les matrices de dispersion inter-classes (S_b^k) et intra -classes (S_w^k)
$$S_b^k = \sum_{y=1}^{\prod_{i \neq k} I_i} S_b^y, S_b^y = \sum_{j=1}^C C_j (\bar{\mathbf{z}}_j^{k,y} - \bar{\mathbf{z}}^{k,y}) (\bar{\mathbf{z}}_j^{k,y} - \bar{\mathbf{z}}^{k,y})^T$$

$$S_w^k = \sum_{y=1}^{\prod_{i \neq k} I_i} S_w^y, S_w^y = \sum_{i=1}^M (\mathbf{z}_i^{k,y} - \bar{\mathbf{z}}_c^{k,y}) (\mathbf{z}_i^{k,y} - \bar{\mathbf{z}}_c^{k,y})^T$$
 - ♦ Calculer la nouvelle matrice de dispersion inter-classes (N_b^k)
$$S_w^k = (V_w^k)^T \Lambda_w^k V_w^k$$

$$N_b^k = (V_w^k)^{-1/2} (\Lambda_w^k)^T S_b^k \Lambda_w^k (V_w^k)^{-1/2}$$
 - ♦ Calculer la matrice de projection finale Y_k^{itr}

$$N_b^k = (E^k)^T O_b^k E^k$$

$$Y_k^{\text{itr}} = \Lambda_w^k V_w^k E^k, Y_k^{\text{itr}} \in \mathbb{R}^{I_k \times I'_k}$$
 - b) **Si** itr > 2 et $\|Y_k^{\text{itr}} - Y_k^{\text{itr}-1}\| < I_k \times I_k \times \varepsilon, k = 1, \dots, L$, pause ;
3. Calculer les dimensions inférieures finales I'_k par: $\text{Energie}_k = \frac{\sum_{i=1}^{I'_k} \lambda_i^k}{\sum_{i=1}^{I_k} \lambda_i^k} \times 100$
4. Trier I'_k vecteurs propre $Y_k^{\text{itr}} \in \mathbb{R}^{I_k \times I'_k}$ en fonction de λ_i^k par ordre décroissant, $K = 1, \dots, L$.

4.3 Aperçu de l'approche de vérification de visage proposée

Dans cette section, nous expliquons les détails de l'approche proposée pour la vérification de visage multimodale 2D et 3D basée sur la représentation du tenseur. Comme illustré à la figure. 4.1, le schéma fonctionnel du système proposé comprend trois composants essentiels : conception du tenseur basé sur les caractéristiques locales, réduction de la dimensionnalité et la classification.

4.3.1 Conception du tenseur basé sur les caractéristiques locales

Les données Tenseur du visage sont construites en utilisant des caractéristiques locales sur trois étapes; le prétraitement des images 2D et 3D, l'extraction de caractéristiques basé sur le descripteur locale LPQ (Local Phase Quantization) [130] et la représentation du tenseur. La procédure d'extraction des caractéristiques et la conception du tenseur est illustrée dans la figure 4.2. Premièrement, le prétraitement des visages comprend plusieurs étapes pour améliorer la qualité des données 2D et 3D. La zone du visage est recadrée à l'aide d'un masque elliptique centré sur le bout du nez. La zone du visage est découpée à l'aide d'un masque elliptique centré sur le bout du nez. Les coordonnées des points de bout du nez sont localisées par la méthode de l'intensité maximale [126, 131], la région faciale est découpée à partir de l'image profondeur 3D et de son image d'intensité 2D correspondante. Généralement, les données obtenues à partir du processus de numérisation 3D sont bruyantes, contient également différents trous dans certaines parties. Pour cela, un filtre médian est appliqué afin de supprimer le bruit. Aussi, nous utilisons également une interpolation cubique des pixels voisins pour remplir les trous.

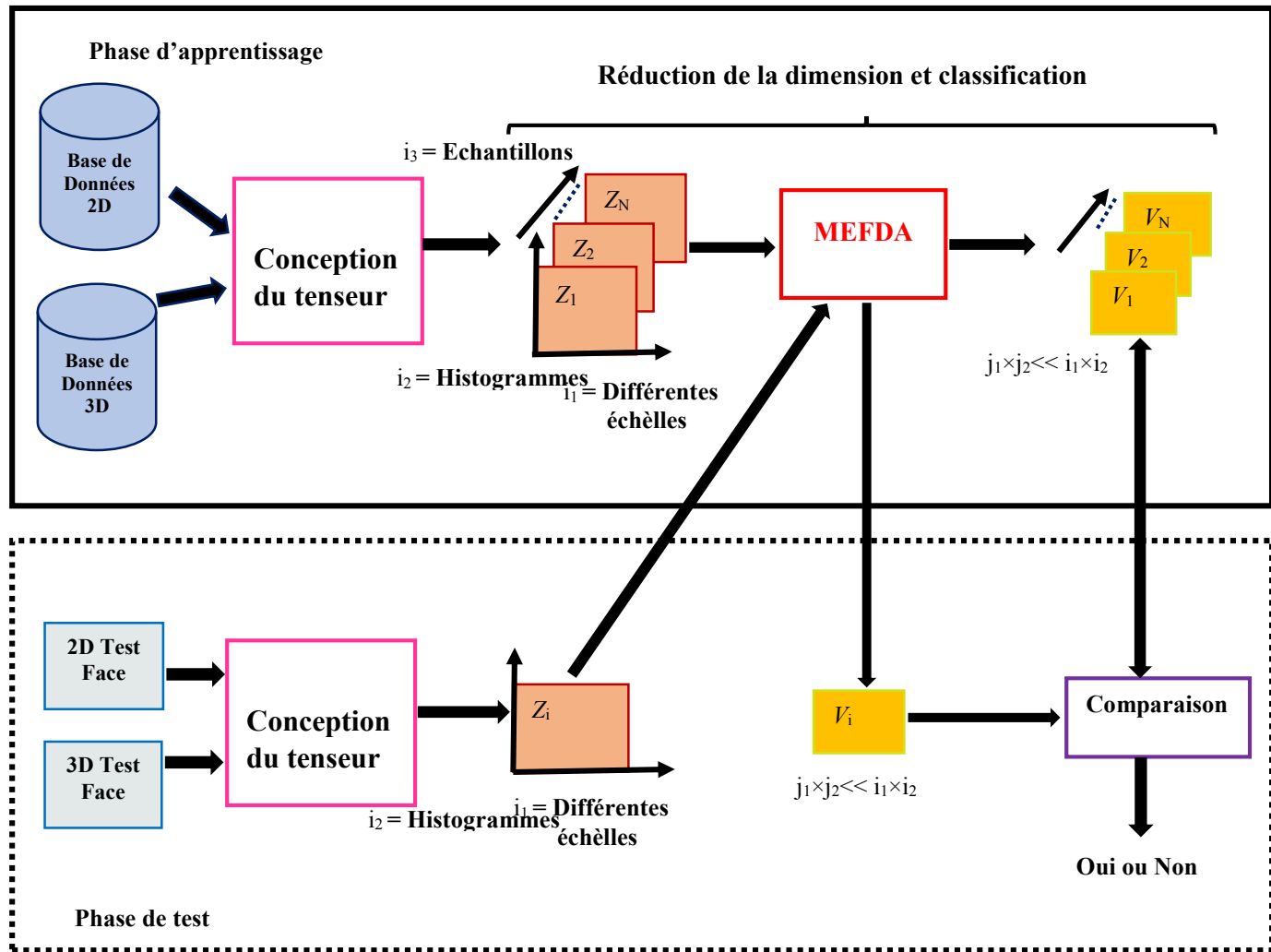


Figure 4.1: Système proposé pour la vérification multimodale de visage 2D+3D.

Ensuite, une égalisation d'histogramme est appliquée pour réduire l'influence des variations d'illumination dans les images d'intensité 2D.

Deuxièmement, pour l'extraction de caractéristiques, les images de profondeur et d'intensité sont décrites en utilisant le descripteur LPQ. Nous utilisons ce descripteur car il a réalisé les meilleures performances dans des travaux précédents [131] [132, 133]

4.3.1.1 Descripteur LPQ (Local Phase Quantization)

Le descripteur de texture LPQ a été proposé par Ojansivu et Heikkilä [130], Il permet d'améliorer la classification de textures afin d'être robuste aux artéfacts générés par le flou présent dans une image [134]. Le descripteur LPQ est construit de façon à ne garder dans une image que l'information locale invariante à un certain type de flou qui sera symétrique, qui causé généralement par le mouvement linéaire et hors du foyer du capteur [130]. Inspiré par cette idée, nous proposons le descripteur LPQ comme une méthode efficace pour résoudre le

problème des variations d'expressions dans le système de vérification du visage multimodale 2D+3D.

L'opérateur LPQ est basé sur la transformation de Fourier de la phase d'où on utilise cette dernière pour l'extraction de l'information de la phase locale basée sur l'application de la transformée de Fourier à court terme (STFT) calculé sur une zone rectangulaire M_x de $N * N$ voisins à chaque position de pixel x de l'image $f(x)$ définie par

$$F(u, x) = \sum_{v \in x} f(x - y) e^{2j\pi u^T y} = w_u f_x \quad (4.18)$$

où w_u correspond aux vecteurs de base de la décomposition à la fréquence u , f_x contient toutes les valeurs de l'image appartenant au voisinage N_x . Dans LPQ seulement quatre coefficients complexes sont considérés $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$ et $u_4 = [a, -a]^T$, où a est une scalaire suffisamment faible pour satisfaire la condition $H(u_i) > 0$. Pour chaque position de pixel cela se traduit par un vecteur f_x^c

$$f_x^c = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)] \quad (4.19)$$

L'information de phase dans les coefficients de Fourier est enregistrée, en observant les signes des parties réelles et imaginaires de chaque composante en $F(x)$. Ceci est fait en utilisant une quantification scalaire simple.

$$q_i(x) = \begin{cases} 1, & \text{si } g_j(x) \geq 0 \\ 0, & \text{sinon} \end{cases} \quad (4.20)$$

Où g_j représente la $j^{\text{ième}}$ composante du vecteur $g(x) = [\text{Re } \{F(x)\}, \{\text{Im } F(x)\}]$. Les huit coefficients binaires $q_i(x)$ résultants sont représentés par des valeurs entières entre 0 à 255 en utilisant un codage binaire simple pour obtenir les étiquettes de LPQ, F_{LPQ} qui est définie par

$$F_{LPQ}(X) = \sum_{i=1}^8 q_i(x) 2^{j-1} \quad (4.21)$$

La figure 4.3 présente l'organigramme de l'ensemble des étapes nécessaire à la construction du descripteur LPQ pour une image visage.

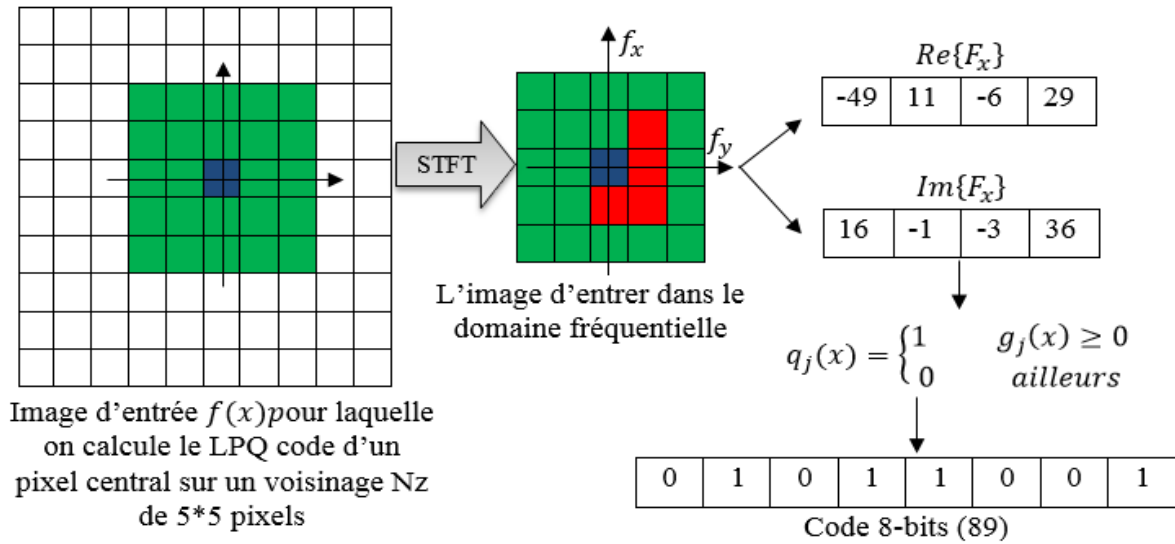


Figure 4.2: Organigramme de l'ensemble des étapes nécessaires du descripteur LPQ.

Afin de conserver la structure spatiale des visages, les images de visage 2D et 3D sont subdivisées sans chevauchement en P régions de visages dans lesquels les histogrammes de ces blocs rectangulaires sont concaténés pour former un vecteur de caractéristiques v de taille $n = P \times 256$. Pour améliorer encore la description des visages, nous utilisons la multi-échelle LPQ (MSLPQ) pour l'extraction de caractéristiques des images 2D et 3D. Notez que LPQ est en fait un cas particulier de MLPQ avec une échelle, la représentation multi-échelle LPQ peut être obtenue en faisant varier la taille de la fenêtre M . Par conséquent, les caractéristiques LPQ sont extraites à différentes échelles. Les vecteurs caractéristiques de différentes échelles sont arrangés dans une matrice formant une représentation du tenseur du second ordre. Finalement, tous les candidats de visage sont arrangés pour créer le tenseur de visage de troisième ordre $X \in \mathbb{R}^{i_1 \times i_2 \times i_3}$.

Par conséquent, les trois modes du tenseur sont définis comme suit:

i_1 : Les vecteurs caractéristiques à différentes échelles utilisées pour les images de visage 2D et 3D.

i_2 : Les histogrammes concaténés.

i_3 : Les différentes personnes de la base de données.

4.3.2 Réduction de la dimensionnalité et classification

Dans cette étape, les données de tenseur d'apprentissage $\mathbf{Z} \in \mathbb{R}^{i_1 \times i_2 \times i_3}$ sont projetées dans un sous-espace discriminant inférieur. Dans la phase d'apprentissage, les matrices de projection multilinéaires optimales sont estimées, et dans la phase de test, les nouveaux échantillons sont projetés à l'aide de ces matrices optimales et comparés. Premièrement, la MPCA est appliquée avant la MEFDA proposée afin d'éliminer le problème de petite taille d'échantillon dans différents modes tensoriels. Ce problème se produit lorsque le nombre de caractéristiques est supérieur au nombre d'échantillons. Comme le troisième mode du tenseur i_3 représente les différentes personnes de la base de données mentionnées ci-dessus, nous ne pouvons pas réduire notre tenseur selon ce mode. Après l'utilisation de la MPCA, les données de tenseur d'apprentissage sont projetées dans un nouveau sous-espace basé sur notre méthode MEFDA selon leurs deux modes i_1 et i_2 (échelles et caractéristiques). Par conséquent, un nouveau tenseur réduit et discriminant $\mathbf{V} \in \mathbb{R}^{j_1 \times j_2 \times i_3}$ est obtenu, où $j_1 \times j_2 \ll i_1 \times i_2$.

Pendant la phase de test, l'image de visage de test pour chaque candidat Z_i de taille $i_1 \times i_2$ est projetée dans le sous-espace MEFDA avec la même procédure que celle utilisée lors de la phase d'apprentissage. Ensuite, nous obtenons des caractéristiques réduites $V_i \in \mathbb{R}^{j_1 \times j_2}$, où $j_1 \times j_2 \ll i_1 \times i_2$. Finalement, pour la classification, les caractéristiques de test réduites V_i sont comparées aux données de tenseur d'apprentissage \mathbf{V} en utilisant la comparaison basée sur la similarité de cosinus.

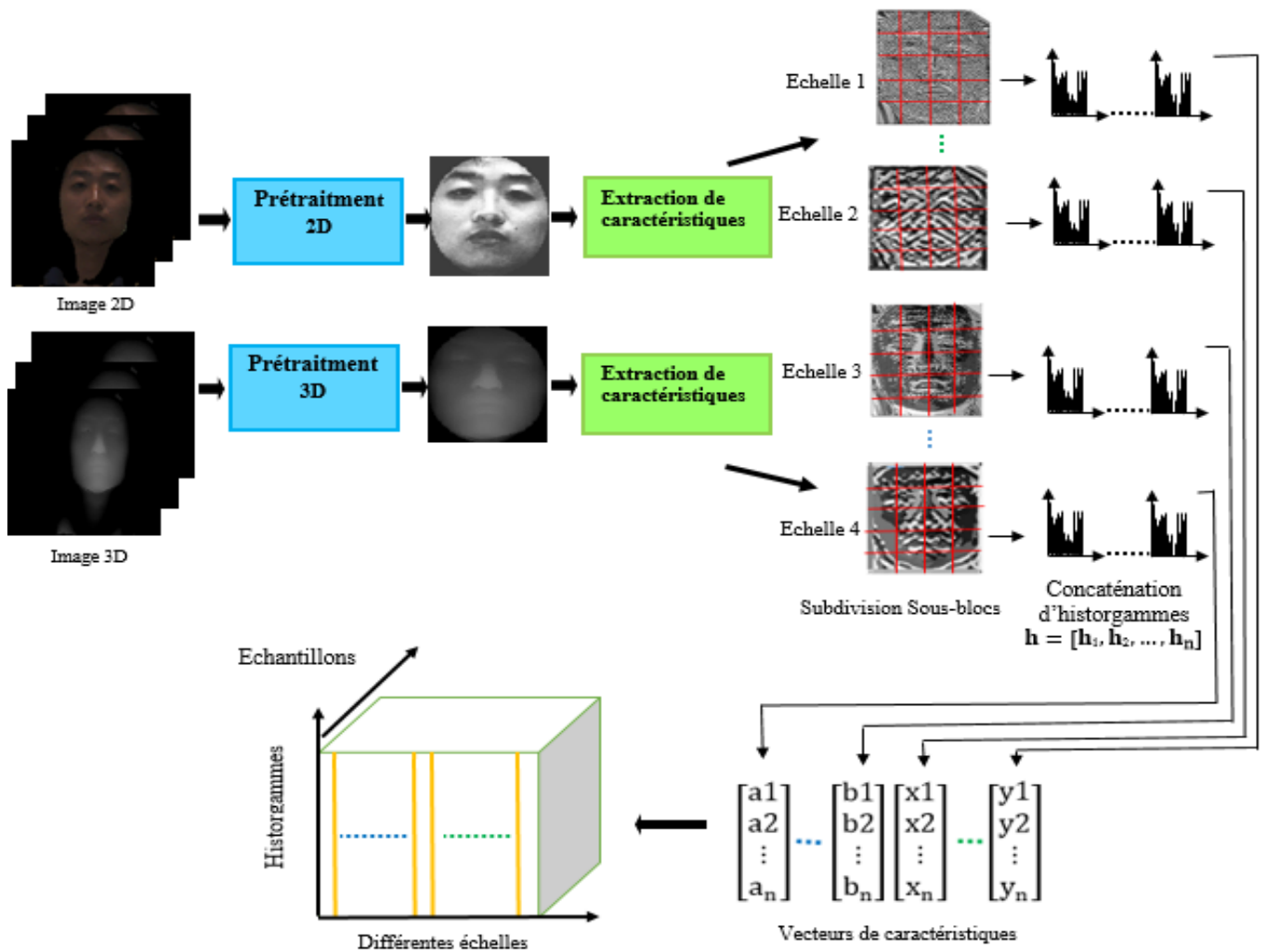


Figure 4.3: Extraction des caractéristiques et conception du tenseur.

4.4 Evaluation expérimentale

Dans cette section, nous évaluons l'approche proposée sous plusieurs expériences pour la vérification du visage basée sur la représentation du tenseur du 3^{ème} ordre. Premièrement, nous décrivons les trois bases de données de visage populaires, FRGC V2.0 [135], CASIA 3D [136] et Bosphorus [137], utilisées dans nos expériences, aussi bien que, nous donnons les paramètres de réglage en détail. Ensuite, nous procédons à l'évaluation de la vérification de visage uni-modale 2D, uni-modale 3D et multimodale 2D+3D. Finalement, nos meilleurs résultats sont comparés avec l'état de l'art.

4.4.1 Base de données de visage

4.4.1.1 Base de données FRGC V2.0

FRGC V2.0 est l'une des plus grande actuellement base de données disponible de visage 3D. Elle est constituée de 4950 scans utilisant le scanner laser Minolta Vivid 900/910. La base de données offre de nombreux défis: haute diverses expressions, différentes conditions d'illuminations et plus de 40% des sujets ont des occlusions différentes. Pour les expériences avec la base FRGC, nous évaluons l'approche proposée sur le même protocole de vérification que celui défini dans [138], où les données sont divisées en deux partitions: apprentissage et validation. L'ensemble d'apprentissage comprend 943 visages de 266 personnes, tandis que l'ensemble de validation comprend 4007 visages de 466 personnes. Des exemples d'images de visage de cette base de données sont présentés à la figure 4.4.



Figure 4.4: Exemples d'images de visage de la base de données FRGC V2.0.

4.4.1.2 Base de données CASIA 3D

La base de données CASIA de visages 3D comprenant 4624 scans de 123 sujets utilisant le scanner 3D Minolta Vivid 910 sans contact fonctionnant en mode rapide. Chaque sujet a 37 ou 38 images avec des variations individuelles d'expressions, de poses, d'illuminations, changements combinés dans les expressions sous les poses et les illuminations. Dans nos expériences, nous évaluons l'approche proposée en utilisant le même protocole de test que celui défini dans [117, 131, 139], dans lesquels 15 modèles différents pour chaque sujet sont utilisés comme suit: variations d'illuminations (scans: 1 à 5), variations d'expression (scans: 6 à 10) et expressions sous les variations d'illuminations (scans: 11 à 15).

Le protocole sépare les personnes en deux classes, client et imposteur. Le groupe de clients comprend 100 personnes, tandis que le groupe d'imposteurs est divisé en 13 personnes pour l'ensemble d'évaluation et 10 personnes pour l'ensemble de tests. Des exemples d'images de visage de cette base de données sont présentés à la figure 4.5.



Figure 4.5: Exemples d'images de visage de la base de données CASIA 3D.

4.4.1.3 Base de données Bosphorus

La base de données de visages Bosphorus comprend 4666 numérisations de 105 sujets différents dans diverses expressions, poses et occlusions. La base de données Bosphorus est collectée par un scanner 3D à lumière structurée. Dans nos expériences, nous évaluons l'approche proposée sur le même protocole que celui défini dans [140], dans lequel un total de 1097 visages de 105 sujets sont utilisés. La phase d'apprentissage contient 580 images de 58 sujets, 47 images avec une expression neutre sont utilisées pour une galerie et la phase de test contient 470 images de 47 sujets. Des exemples d'images de visage de cette base de données sont présentés à la figure 4.6.



Figure 4.6: Exemples d'images de visage de la base de données Bosphorus.

4.4.2 Paramètres de réglage

Les paramètres de réglage sont choisis expérimentalement pour donner les meilleurs résultats. Les images de visage sont découpées et redimensionnées à 150×130 pixels. Dans l'étape d'extraction des caractéristiques, le descripteur LPQ multi-échelle (MSLPQ) est appliqué en utilisant quatre échelles $\{5, 9, 13, 17\}$. Ensuite, chaque image de visage est subdivisée en 25 blocs sans chevauchement, où la taille de chaque bloc est de 30×30 pixels. Les histogrammes avec 256 éléments de ces blocs sont extraits et concaténés en un vecteur de caractéristiques de

taille 25×256 . Le nombre d'itérations pour l'algorithme MEFDA est réglé de manière empirique dans la meilleure valeur dans laquelle $\text{Iteration}_{\text{Max}} = 12$, et la dernière dimension inférieure de sous-espace est défini à 99% de l'énergie des valeurs propres. L'estimation des matrices de transformation de sous-espace est effectuée à l'aide des données d'apprentissage et la dimension inférieure optimale des vecteurs de caractéristiques est sélectionnée. Afin d'évaluer les performances de vérification de différentes approches avec les bases de données FRGC V2.0 et Bosphorus, nous rapportons le taux de vérification (TV) à 0,001 TFA (Taux de Faux Acceptations). Pour CASIA 3D, nous rapportons le taux d'égale erreur (TEE) de l'ensemble d'évaluation et le taux de vérification au seuil correspondant de l'ensemble de test. Ainsi, toutes les courbes ROC (Receiver Operating Characteristic) sont tracés afin de comparer les différentes méthodes.

4.4.3 Résultats et discussion

Dans cette sous-section, nous présentons différentes évaluations pour montrer l'efficacité de l'approche proposée. En termes de comparaison, les performances de notre algorithme MEFDA sont comparées à un certain nombre de méthodes basiques de projection de sous-espace. Pour les algorithmes linéaires, nous évaluons PCA, LDA et EFM, qui servent de base à l'évaluation du MEFDA proposé. Pour les algorithmes multilinéaires basés sur les tenseurs du 3ème ordre, nous évaluons les méthodes de base MPCA et MDA.

Pour supprimer le problème de petite taille d'échantillon dans le cas de la représentation linéaire, La PCA est appliquée avant LDA et EFM et, dans le cas de la représentation multilinéaire, la MPCA est effectuée avant l'application de MDA et MEFDA. Les résultats obtenus de la précision de la vérification de visage dans chaque base de données sont fournis dans le tableau 4.1 pour FRGC V2.0. Tableau 4.2 pour CASIA 3D ainsi que dans le tableau 4.3 pour la base de données Bosphorus. L'analyse des résultats obtenus est donnée ci-dessous.

4.4.3.1 Vérification uni-modale de visage 2D et 3D

En comparant les performances de 2D contre 3D sur les trois bases de données, pour chaque cas particulier des tableaux 4.1, 4.2 et 4.3, les systèmes travaillant sur les données de visage 3D sont toujours meilleurs que ceux obtenus par des informations 2D. La différence de taux de vérification est plus évidente et remarquable, où les meilleurs taux de vérification utilisant des données de visage 3D sont de 96,35%, 99,06% et 95,31% sur les bases de données FRGC V2.0, CASIA 3D et Bosphorus, respectivement. De plus, le TEE obtenu est inférieur pour les données 3D par rapport aux données 2D. Par conséquent, nous pouvons conclure que le pouvoir

discriminant des images de visage 3D avec différents défis, tels que des illuminations, des poses et des expressions, est supérieur à celui fourni par les images 2D.

4.4.3.2 Vérification multimodale de visage 2D + 3D

Les performances de vérification présentés dans les tableaux 4.1, 4.2 et 4.3 montrent que la fusion de images de visage (2D + 3D) donne de meilleurs résultats que les images 2D et 3D uni-modale séparément. L'augmentation des performances est plus remarquable sur les bases de données FRGC V2.0 et CASIA 3D, où les taux de vérification atteignent une augmentation de 2% à 4%. Par conséquent, il est clair que la fusion entre les modalités 2D et 3D fournit plus de descriptivité aux informations de visage et prouve clairement la complémentarité entre elles. En outre, les résultats obtenus confirment que notre stratégie de fusion des informations 2D et 3D basées sur le tenseur d'ordre supérieur est supérieure à la fusion basée sur la représentation matricielle de caractéristiques. Les meilleurs taux de vérification sont obtenus avec les modes multimodaux 2D + 3D, dans lesquels nous avons 99.00%, 99.36% et 97.02% sur les bases de données FRGC V2.0, CASIA 3D et Bosphorus, respectivement.

Tableau 4.1: Taux de vérification (%) à 0,001 TFA uni-modale 2D, 3D et multimodal 2D + 3D sur la base FRGC V2.0.

Méthode	Image intensité 2D				Image de profondeur 3D				2D+3D			
	ROC I	ROC II	ROC III	N vs. all	ROC I	ROC II	ROC III	N vs. all	ROC I	ROC II	ROC III	N vs. all
PCA [121]	55.98	46.32	34.93	53.73	64.14	60.47	56.72	68.48	81.50	80.21	78.70	82.75
LDA [71]	68.51	58.48	47.11	64.21	79.79	77.71	75.52	80.85	88.30	87.42	86.20	88.79
EFM [129]	69.53	58.83	46.93	64.23	88.43	87.51	86.43	88.64	91.79	91.10	90.36	92.01
MPCA [24]	75.71	64.54	52.58	69.40	84.68	82.86	80.55	87.42	91.68	84.32	76.05	87.27
MDA [125]	86.30	76.33	66.43	80.80	92.14	90.94	89.68	93.91	95.80	91.80	87.52	94.08
MEFDA	87.04	77.03	67.03	82.08	95.94	95.25	94.40	96.35	99.00	98.34	97.81	98.22

Tableau 4.2: Performances de la vérification (%) uni-modale 2D, 3D et multimodales 2D + 3D sur la base CASIA 3D.

Méthode	Image intensité 2D		Image de profondeur 3D		2D+3D	
	Evaluation	Test	Evaluation	Test	Evaluation	Test
PCA [121]	EER (%)	VR (%)	EER (%)	VR (%)	EER (%)	VR (%)
LDA [71]	5.07	89.24	4.96	90.89	2.75	93.27
EFM [129]	4.82	89.66	2.47	95.69	1.15	97.14
MPCA [24]	3.36	90.35	1.67	96.62	0.59	97.38
MDA [125]	3.17	91.86	5.61	93.54	1.76	97.63
MEFDA	1.57	94.45	1.14	96.64	0.60	98.59

Tableau 4.3: Taux de vérification (%) à 0,001 TFA uni-modale 2D, 3D de visage et multimodal 2D + 3D sur la base Bosphorus.

Méthode	Image intensité 2D	Image de profondeur 3D	2D+3D
PCA [121]	43.40	79.57	59.14
LDA [71]	58.72	83.40	75.53
EFM [129]	87.65	89.14	92.76
MPCA [24]	47.44	87.02	79.78
MDA [125]	61.06	91.06	84.46
MEFDA	88.93	95.31	97.02

4.4.3.3 MEFDA vis à vis EFM

Dans cette expérience, nous comparons la MEFDA proposée pour le tenseur d'ordre élevé à l'algorithme de contrepartie de base EFM. Le taux de vérification avec les différentes modalités 2D, 3D et 2D + 3D sur les trois bases de données est illustré à la figure 4.7. La vision profonde des résultats prouve et confirme que MEFDA est supérieure à leur contrepartie linéaire EFM. Certainement, MEFDA est capable d'extraire plus de sous-espaces de caractéristiques discriminants que l'EFM en profitent pleinement l'avantage de la représentation du tenseur d'ordre élevé. Dans le cas linéaire EFM, les caractéristiques sont simplement concaténées en ignorant la structure de données naturelle. Sachant que, dans le cas multilinéaire utilisant MEFDA, les données de visage sont empilées en tant qu'entrée au niveau multicanal à l'aide d'un tenseur d'ordre élevé, ce qui permet d'extraire une signature de visage naturelle. Par ailleurs, MEFDA domine mieux le contenu de pouvoir discriminant existant dans chaque mode k du tenseur par rapport à EFM.

En prenant comme exemple l'utilisation de données de visage (2D + 3D) dans chaque base de données, les taux de vérification dans CASIA 3D nous donnent $TV = 96,62\%$ avec EFM et $TV = 99,06\%$ avec notre méthode MEFDA. Pour la base de données Bosphorus, nous obtenons $TV = 89,14\%$ et $TV = 95,31\%$ pour EFM et MEFDA respectivement. De même, pour FRGC V2.0 dans le test ROC I, nous obtenons $TV = 91,79\%$ et $TV = 99\%$ pour EFM et MEFDA

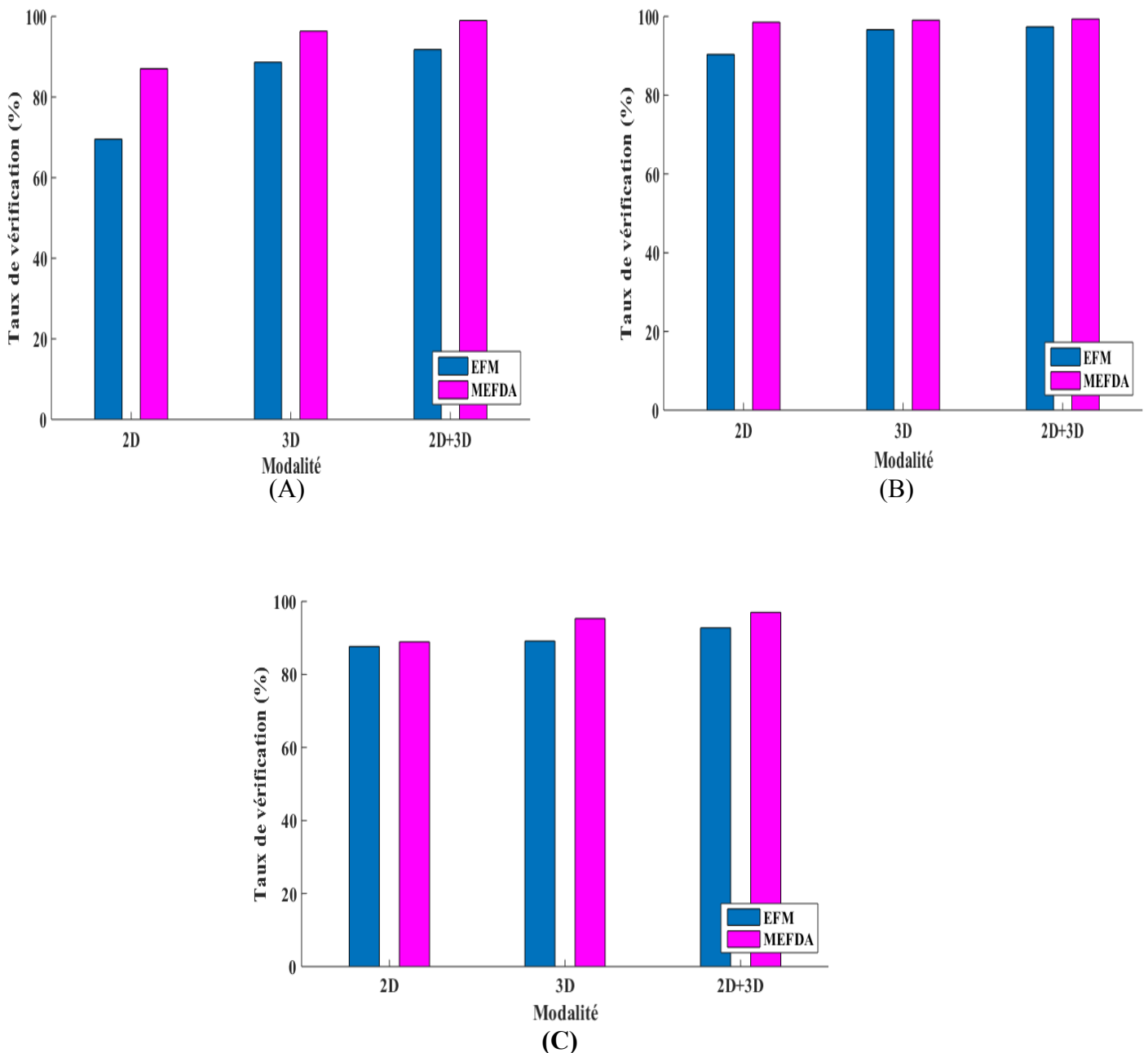


Figure 4.7: Taux de vérification de EFM et MEFDA pour 2D, 3D et 2D + 3D sur (a) FRGC V2.0 (b) CASIA 3D (c) Bosphorus, respectivement.

4.4.3.4 Le temps de complexité

Les expériences ont été réalisées sous MATLAB version 2015a sur un PC avec un processeur Intel Core i3 à 1,80 GHz et de 4,00 Go de RAM. L'estimation des matrices de projection est implémentée dans la phase d'apprentissage hors ligne. La complexité en termes de temps de calcul (TC) pour une correspondance de visage de test pendant la phase de test en ligne en utilisant EFM et la méthode MEFDA proposée est présenté dans le tableau 4.4, Puisque la même taille et les mêmes échelles de descripteur MSLPQ sont appliquées dans toutes les bases de données pour EFM et MEFDA, le même TC du prétraitement de visage et l'extraction de caractéristiques est requis pour les deux méthodes dans chaque base de données. Pour la réduction de la dimensionnalité et la classification, on peut voir que notre méthode MEFDA réalise un bon gain en TC par rapport à la méthode de base dans tous les bases de données. Le TC est réduit grâce à l'utilisation de la représentation du tenseur des données de visage. En plus de la précision, cette expérience démontre l'efficacité de notre approche en termes de complexité de calcul. En effet, le temps total d'exécution avec MEFDA ne dépasse pas un TC total de 1,420 s, ce qui en fait un candidat compétent pour une utilisation dans des applications de vérification du visage dans le monde réel.

Tableau 4.4: Temps de complexité en secondes d'EFM et MEFDA.

Prétraitement du visage	Extraction de caractéristiques	Réduction de la dimensionnalité et la classification		Temps total d'exécution	
		EFM	MEFDA	EFM	MEFDA
0.501	0.633	0.324	0.286	1.458	1.420

4.4.3.5 MEFDA vis à vis toutes les approches

Pour donner plus de crédibilité à notre méthode du tenseur d'ordre élevé, nous la comparons avec différents algorithmes de réduction de dimensionnalité de l'état de l'art, y compris les méthodes PCA [121], LDA [71], MPCA [24] et MDA [125]. Considérant l'influence de toutes les approches de sous-espace, pour chaque cas particulier des tableaux 4.1, 4.2, 4.3, l'approche multilinéaire proposée MEFDA offre les meilleures performances et effectue mieux que les approches linéaires et leurs contreparties multilinéaires comprennent PCA, LDA, EFM, MPCA et MDA. Cela démontre clairement le pouvoir élevé de séparation de MEFDA sous différents facteurs de variabilité (poses, illuminations et variations d'expressions) pour les cas uni-modal (2D et 3D) ainsi que pour leur fusion (2D + 3D) sur les trois bases de données.

En revanche, la comparaison de nos résultats en termes de courbes ROC illustrées aux figures 4.8, 4.9 et 4.10 confirme la supériorité de notre méthode MEFDA par rapport aux autres méthodes, dans lesquelles les taux de vérification utilisant MEFDA à chaque taux de fausse acceptation sont supérieurs au taux de vérification des autres approches, ainsi que les courbes ROC de MEFDA sont plus stables.

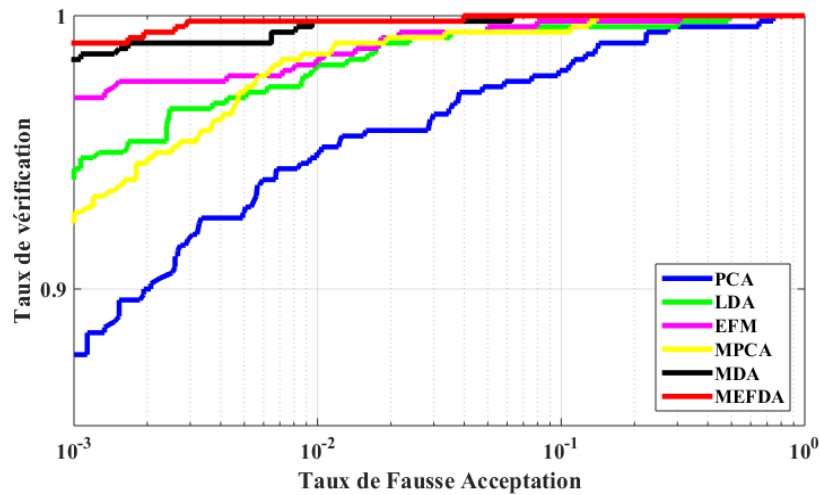


Figure 4.8: Courbe ROC multimodale pour différentes transformations de sous-espaces sur la base de données CASIA 3D.

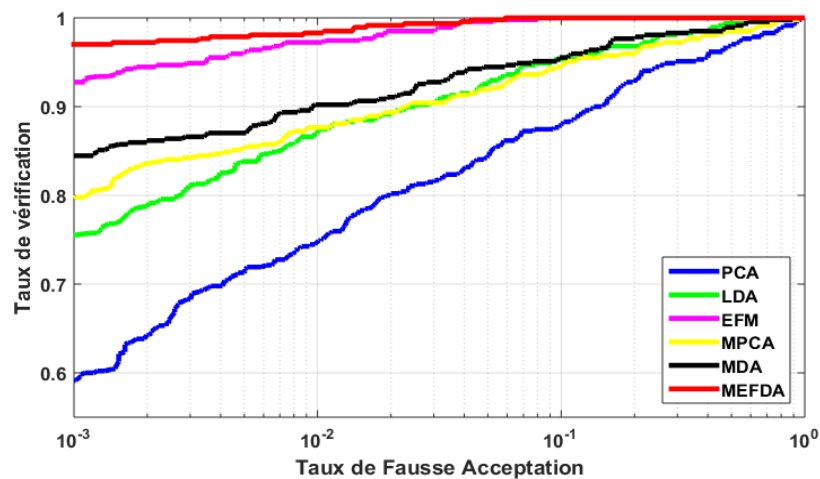


Figure 4.9: Courbe ROC multimodale pour différentes transformations de sous-espaces sur la base de données Bosphorus.

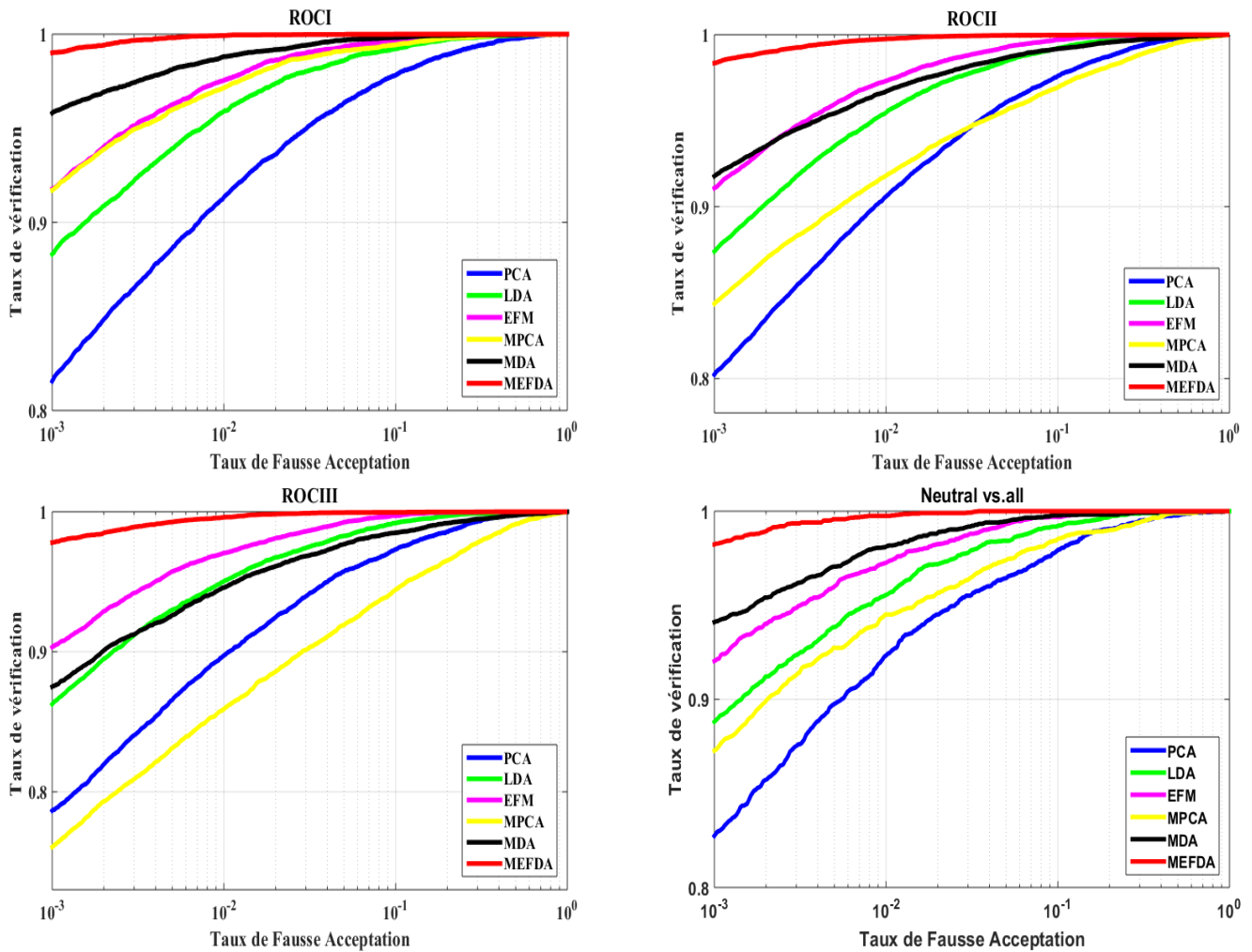


Figure 4.10: Courbes ROC multimodale pour différentes transformations de sous-espaces sur la base de données FRGC V2.0.

4.4.3.6 Comparaison avec l'état de l'art

La comparaison de la MEFDA proposée avec de nombreux algorithmes de l'état de l'art développés au cours des dernières années est présentée dans les tableaux 4.4, 4.5 et 4.6 pour les bases de données FRGC V2.0, CASIA 3D et Bosphorus, respectivement. Pour la première base de données, FRGC V2.0, notre méthode est comparée avec 9 méthodes : [117, 139, 141-146]. Pour la deuxième base de données, CASIA 3D, elle est comparée avec 5 méthodes : [147], [139], [117], [118], [146] et pour la dernière base de données, Bosphorus, elle est comparée avec 6 méthodes : [147], [140], [143], [29], [117], [146]. Le symbole (*) dans les tableaux 4.4 et 4.6 indique que la valeur est extraite de la courbe ROC. Nous notons que les meilleurs taux de vérification de notre travail ont atteint 99% en FRGC V2.0 3D, 99,36% en CASIA 3D et 97,02% en Bosphorus via la combinaison de visage 2D et 3D. Ces résultats sont plus compétitifs

avec l'état de l'art dans la base de données CASIA 3D. Le travail récent dans [117] qui atteint le meilleur taux de vérification actuellement dans le cas multimodal 2D+3D sur les bases de données CASIA 3D et FRGC V2.0 est obtenu par la fusion de deux descripteurs multi-échelles (MSLPQ+MSBSIF) ainsi que par l'application d'une méthode de transformation du sous-espace avec noyau. En revanche, nous avons utilisé un seul descripteur MSLPQ où nous avons atteint le deuxième rang, alors que nous avons atteint le taux de vérification le plus élevé dans l'expérience ROCI sur FRGC V2.0 par rapport aux travaux de la littérature.

En outre, notre approche surpasse les autres méthodes de l'état de l'art dans base de données Bosphorus avec une amélioration remarquable dans laquelle nous obtenons 2,93% avec l'information 2D et 1,63% avec l'information 3D et 1,12% sur leur fusion 2D + 3D. Généralement, la comparaison avec les méthodes de l'état de l'art démontre bien l'efficacité de notre méthode.

Tableau 4.4: Comparaison de taux de vérification (%) avec l'état de l'art sur la base de données FRGC V2.0.

Auteurs	Méthode	Modalité	ROC I	ROC II	ROC III	N vs. all
Lei et al [4]	ARS+ KPCA	3D	/	/	96.70	/
Al-Osaimi* [29]	RAIKs		95.80	96.50	96.60	/
Emambakhsh et Evans [31]	Spherical patches		/	/	93.50	/
Ouamane et al [139]	EDA+ WCCN		92.28	91.25	90.29	93.83
Ouamane et al [117]	TEDA+WCCN		96.98	95.93	95.10	96.65
Notre méthode	MEFDA		95.94	95.25	94.40	96.35
Elaiwat et al [143]	Curvelet local features	2D + 3D	/	/	97.80	/
Soltanpour et Jonathan* [144]	pyramidal shape map		/	/	95.00	98.20
Ioannis et al* [145]	UR2D-R		/	/	72.00	/
Ouamane et al [139]	EDA+ WCCN		97.46	95.42	93.93	97.87
Ouamane et al [117]	TEDA+ WCCN		98.95	98.50	98.44	98.32
Guo et al [146]	LGP-WHGO		/	/	98.06	/
Notre méthode	MEFDA		99.00	98.34	97.81	98.22

Tableau 4.5: Comparaison de taux de vérification (%) avec l'état de l'art sur la base de données CASIA 3D.

Auteurs	Méthode	3D	2D+3D
Ming [147]	ROSR	96.25	/
Ouamane et al [139]	EDA+WCCN	96.97	97.01
Ouamane et al [117]	TEDA+WCCN	98.97	99.47
Ammar et al [118]	HOSVD+EFM	98.60	/
Guo et al [146]	LGP-WHGO	/	98.61
Notre méthode	MEFDA	99.06	99.36

Tableau 4.6: Comparaison de taux de vérification (%) avec l'état de l'art sur la base de données Bosphorus.

Auteurs	Méthode	3D	2D+3D
Ming [147]	ROSR	93.95	/
Erdogmus et al [140]	LBP	/	69.09
Elaiwat et al [143]	Curvelet local features	77.50	91.00
Al-Osaimi* [142]	RAIKs	91.00	/
Ouamane et al [117]	TEDA+WCCN	92.97	96.17
Guo et al [146]	LGP-WHGO	/	96.75
Notre méthode	MEFDA	95.31	97.02

4.5 Conclusion

Dans ce chapitre, nous avons présenté une approche de vérification multimodale de visage (2D + 3D) robuste dans des conditions incontrôlables telles que des poses, des expressions et des occlusions. Sur un descripteur LPQ basé sur des sous-blocs multi-échelles, les caractéristiques locales sont extraites et représentées sous la forme d'un tenseur de troisième ordre. Afin de projeter les échantillons de tenseurs d'entrée dans un sous-espace de faible dimension et d'améliorer la discrimination entre différentes images de visage, une nouvelle méthode de transformation de sous-espace multilinéaire appelée Multilinear Enhanced Fisher Discriminant Analysis (MEFDA) est proposée. MEFDA trouve une matrice de projection optimale dans chaque mode tenseur et extrait des informations plus discriminantes, ce qui améliore la séparation entre les classes de données. Notre approche a été validée sur trois bases de données

Chapitre 4 MEFDA pour la vérification multimodale du visage 2D+3D robuste

publiques FRGC v2.0, CASIA 3D et Bosphorus. La comparaison avec les méthodes de l'état de l'art a montré que notre nouvelle méthode MEFDA surpasse les approches linéaires classiques ainsi que leurs contreparties multilinéaires MPCA et MDA. Comme une direction future, nous prévoyons d'étudier l'intégration de la transformation multilinéaire de sous-espaces en utilisant des tenseurs d'ordre élevé avec des méthodes d'apprentissage approfondi pour la vérification des visages.

Chapitre 5

MSIDA pour la vérification du visage et de la parenté

5.1 Introduction	97
5.2 Etat de l'art	98
5.3 MSIDA (Multilinear Side-Information based Discriminant Analysis)	99
5.3.1 SILD	99
5.3.2 MSIDA	100
5.4 Vérification des paires de visages basés sur MSIDA	104
5.4.1 Extraction de caractéristiques	105
5.4.1.1 Descripteur LPQ (Local Phase Quantization)	105
5.4.1.2 Descripteur BSIF (Binarized Statistical Image Features)	106
5.4.2 Matching	107
5.5 Expériences	108
5.5.1 Base de données de visage	108
5.5.1.1 Base de données LFW (Labeled Faces in the Wild)	108
5.5.1.2 Base de données Cornell KinFace	109
5.5.1.3 Base de données UB KinFace	109
5.5.1.4 Base de données TSKinFace	110
5.5.2 Paramètres de réglage	111
5.5.3 Résultats et discussion	111
5.5.4 Effet de la quantité de paires d'apprentissage	114
5.5.5 MSIDA faiblement supervisé contre LDA et MDA supervisé	116
5.5.6 Temps de complexité	117
5.5.7 Comparaison avec l'état de l'art	118
5.5.7.1 Vérification de visage dans les environnements non contrôlés	118
5.5.7.2 Vérification de parenté dans les environnements non contrôlés	120
5.6 Conclusion	121

5.1 Introduction

Au cours des dernières années, la reconnaissance automatique du visage dans des environnements contrôlés a atteint des performances satisfaisantes même dans les systèmes biométriques à grande échelle. Cependant, les performances de ces systèmes restent faibles dans des environnements difficiles et non contraintes, caractérisé par qualité faible de l'image, image floue, fortes illuminations de l'environnement et changement d'orientation de la tête, occlusions partielles, etc. Dans ce travail, nous nous concentrons sur le problème de matching (correspondance) des paires d'images du visage [148-150] dans les environnements non contrôlés [46]. En particulier, les applications ciblées consistent à vérifier i) soit une paire d'images faciales sont de la même personne ou non (vérification de l'identité) et ii) si une paire d'images faciales représente deux personnes de la même famille ou non (vérification de la parenté) [151-153]. Dans de telles applications, les données disponibles pour la formation sont faiblement étiquetées. La seule information fournie est de savoir la paire d'images de visage est match (paire positive) ou non-match (paire négative), alors que l'accès à l'identité des personnes n'est pas fourni. Parmi les approches largement utilisées pour la reconnaissance des visages figurent les techniques de transformation de sous-espace [123, 154]. Les approches pionnières de cette catégorie sont les deux méthodes populaires de réduction de la dimensionnalité linéaire, PCA et LDA. Ces algorithmes vectorisent généralement les images de visage, ce qui donne un vecteur de caractéristiques de grande dimension [155]. La vectorisation de l'image provoque également la perte des précieuses informations structurelles sur le visage intégrées aux positions des pixels. Pour résoudre ce problème, des représentations d'ordre élevé des données dans des sous-espaces multilinéaires sont appliquées au problème de reconnaissance des visages afin d'obtenir de meilleures performances [24, 96, 125]. Les approches multilinéaires permettent également une représentation multifactorielle des échantillons d'image de visage, chaque facteur correspondant à une entrée du multi-tableau. Dans ce travail, nous abordons le problème de matching des paires de visage basé sur la représentation du tenseur dans le contexte de données d'apprentissage faiblement étiquetées. Les principales contributions de ce chapitre sont :

1) Nous proposons une nouvelle méthode appelée MSIDA (Multilinear Side-Information based Discriminant Analysis), pour la réduction de la dimensionnalité et la classification des données tensorielles, lorsque l'étiquette de classe de données complète est manquante. MSIDA projette le tenseur de visage d'entrée dans un nouveau sous-espace multilinéaire dans lequel la marge entre échantillons appartenant à des classes différentes est

élargie tandis que la marge dans les échantillons appartenant aux mêmes classes est réduite. De plus, MSIDA réduit la dimension de chaque mode tenseur.

2) Au niveau de la description du visage, nous proposons une nouvelle représentation basée sur les tenseurs d'ordre élevé. Cette représentation combine différents descripteurs locaux, extraits à différentes échelles, offrant une meilleure discrimination. La représentation du tenseur proposée est considérée comme un nouveau moyen pour la fusion des descripteurs locaux.

3) Nous évaluons empiriquement l'approche proposée pour la vérification de l'identité et de la parenté basée sur le visage sur quatre bases de données de visages difficiles (LFW, Cornell KinFace, UB KinFace et TSkinFace). La comparaison avec les méthodes de l'état de l'art démontre l'efficacité de notre approche.

Le reste de ce chapitre est organisé comme suit. Dans la section 5.2, nous discutons en brièvement les travaux de l'état de l'art. Dans la section 5.3, nous décrivons la méthode proposée MSIDA (Multilinear Side-Information based Discriminant Analysis). Dans la section 5.4, l'approche de matching du visage est présentée. La section 5.5 détaille l'évaluation expérimentale, les résultats et l'analyse. La conclusion et les travaux futurs sont donnés dans la section 5.6.

5.2 Etat de l'art

Les approches de réduction de dimensionnalité supervisée nécessitent les informations d'étiquette de classe complètes de chaque échantillon de données pour fonctionner. Toutefois, cette condition ne peut pas toujours être remplie, en particulier dans des conditions non contraintes, où les données sont acquises à partir d'Internet. Dans ces cas, nous n'avons généralement accès qu'à des informations limitées, par exemple une paire d'images appartient à la même classe. Une solution alternative à ce problème est donnée par Kan et al [156]. Les auteurs ont proposé une méthode de réduction de dimensionnalité discriminante, nommée SILD (Side-Information based Linear Discriminant Analysis). SILD calcule directement les matrices de dispersion intra-classe et inter-classe en utilisant les informations faibles (étiquettes de paires d'images). Ouamane et al [133] ont proposé SIEDA (Side-Information based Exponential Discriminant Analysis) en combinant les deux méthodes SILD et EDA (Exponential Discriminant Analysis).

Dans ce travail, nous proposons une extension multilinéaire de SILD pour l'analyse du sous-espace de données tensorielles. Nous appelons le nouvel algorithme MSIDA. Cette méthode est utilisée pour résoudre le problème de matching des paires d'images de visage, dans lequel sont caractérisées par des données faiblement étiquetées. Dans notre cadre, l'ensemble des images du visage est représenté par un tenseur du 3^{ème} ordre construit à l'aide d'histogrammes de deux descripteurs de visage locaux, multi-échelle LPQ (Local Phase Quantization) et multi-échelle BSIF (Binarized Statistical Image Features) [157].

5.3 MSIDA (Multilinear Side-Information based Discriminant Analysis)

Cette section présente la méthode proposée MSIDA (Multilinear Side-Information based Discriminant Analysis). Tout d'abord, nous rappelons brièvement la méthode originale SILD (Side-Information based Linear Discriminant Analysis). Ensuite, l'extension proposée MSIDA est présentée avec une formulation mathématique détaillée.

5.3.1 SILD

Dans de nombreuses situations, déterminer les étiquettes de classe complètes exactes des échantillons de données est une tâche difficile. Au lieu de cela, il est plus facile d'obtenir une sorte d'étiquettes faibles des échantillons. Un exemple de ce cas comprend la vérification du visage dans des conditions non contrôlés, où la seule information connue est de savoir les paires d'images de visage appartiennent ou non aux mêmes personnes. Ces étiquettes faibles (match / non match) sont appelées informations faibles. Dans ce cas, le LDA traditionnel ne peut pas fonctionner, car il s'appuie sur l'étiquette de la classe, ici l'identité de la personne, de chaque échantillon pour calculer les matrices de dispersion. Pour résoudre ce problème, SILD [156] exploite efficacement les informations de l'étiquette faible a été proposé. Dans SILD, les paires positives (match) sont utilisées pour calculer la matrice de dispersion intra-classe S_w et les paires négatives (non match) sont utilisées pour calculer la matrice de dispersion inter-classe S_b

Soit $P = \{(\mathbf{u}_i, \mathbf{u}_j) : l(\mathbf{u}_i) = l(\mathbf{u}_j)\}$ l'ensemble des paires d'images match et $Q = \{(\mathbf{u}_a, \mathbf{u}_b) : l(\mathbf{u}_a) \neq l(\mathbf{u}_b)\}$ l'ensemble des paires d'images non match, avec l indique l'information faible. Les matrices de dispersion SILD intra-classe et inter-classe sont estimées comme suit:

$$S_w^{\text{sild}} = \sum_{(\mathbf{u}_i, \mathbf{u}_j) \in P} (\mathbf{u}_i - \mathbf{u}_j) (\mathbf{u}_i - \mathbf{u}_j)^T \quad (5.1)$$

$$S_b^{\text{sild}} = \sum_{(\mathbf{u}_a, \mathbf{u}_b) \in Q} (\mathbf{u}_a - \mathbf{u}_b) (\mathbf{u}_a - \mathbf{u}_b)^T \quad (5.2)$$

Ainsi, le calcul des matrices de dispersion ne nécessite pas de savoir l'étiquette de classe complète de chaque échantillon. Au lieu de cela, les informations faibles sont exploitées pour calculer les deux matrices. De même à LDA, la matrice de projection dans SILD est obtenue en résolvant le problème d'optimisation suivant:

$$H_{\text{opt}}^{\text{sild}} = \arg \max_H \frac{|H^T S_b^{\text{sild}} H|}{|H^T S_w^{\text{sild}} H|} \quad (5.3)$$

5.3.2 MSIDA

Cette partie détaille l'extension proposée de SILD pour fonctionner sur des données multilinéaires. Soit $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ une représentation tensorielle d'échantillons d'apprentissage. L'ensemble d'échantillons d'apprentissage est organisé en paires de tenseurs match $S_p = \{(\mathbf{X}_o, \mathbf{X}_p) : l(\mathbf{X}_o) = l(\mathbf{X}_p)\}$ et l'ensemble des paires de tenseurs non match $D_p = \{(\mathbf{X}_r, \mathbf{X}_s) : l(\mathbf{X}_r) \neq l(\mathbf{X}_s)\}$, où l désigne ici l'informations faible. Afin d'améliorer la discrimination entre les tenseurs de différentes classes, un critère du tenseur discriminant (DTC) est imposé. Dans [125] le DTC garantit que la dispersion des échantillons intra-classe est minimisée et que la dispersion des échantillons inter-classe est maximisée. Différemment, dans notre cas, puisque nous n'avons pas les étiquettes de classes, nous minimisons la dispersion des paires similaires et maximisons la dispersion des paires dissimilaires, comme dans SILD. Le DTC est obtenu par de multiples projections interdépendantes du tenseur à travers ses différents modes. Ainsi, le DTC dans notre cas est donné par:

$$U_{k=1}^*{}^N = \arg \max_{U_{k=1}^N} \frac{\sum_{D_p=1}^{n_{D_p}} \|\mathbf{X}_r \times_1 U_1 \dots \times_N U_N - \mathbf{X}_s \times_1 U_1 \dots \times_N U_N\|^2}{\sum_{S_p=1}^{n_{S_p}} \|\mathbf{X}_o \times_1 U_1 \dots \times_N U_N - \mathbf{X}_p \times_1 U_1 \dots \times_N U_N\|^2} \quad (5.4)$$

n_{S_p} et n_{D_p} sont le nombre de paires positives et de paires négatives, respectivement.

L'équation (5.4) est un problème d'optimisation avec une contrainte non linéaire d'ordre élevé. Par conséquent, trouver sa solution fermée est une tâche difficile. Une solution alternative consiste à utiliser l'approche d'optimisation itérative comme dans [125] pour estimer les matrices de projection interdépendantes.

Tout d'abord, nous considérons un mode du tenseur, l'optimisation est formulée comme suit:

$$U_k^* = \arg \max_{U_k} \frac{\sum_{D_p=1}^{n_{Dp}} c_j \|X_r \times_k U_k - X_s \times_k U_k\|^2}{\sum_{S_p=1}^{n_{Sp}} \|X_o \times_k U_k - X_p \times_k U_k\|^2} \quad (5.5)$$

Comme prouvé dans [125], le problème d'optimisation de l'équation (5.5) est un cas particulier d'analyse discriminante. Pour résoudre ce problème, les tenseurs sont dépliés en matrices dans le mode k . Ensuite, les vecteurs correspondants de la matrice dépliant sont attribués la même étiquette que le tenseur d'origine. Les paires de vecteurs positifs sont utilisés pour calculer la matrice de dispersion intra-classes et les paires de vecteurs négatives sont utilisées pour calculer la matrice de dispersion inter-classes. Ainsi, le problème d'optimisation dans l'équation (5.5) pourrait être réécrit comme suit:

$$U_k^* = \arg \max_{U_k} \frac{\text{Tr}(U_k^T (S^{\text{msida}}_b)^k U_k)}{\text{Tr}(U_k^T (S^{\text{msida}}_w)^k U_k)} \quad (5.6)$$

où $(S^{\text{msida}}_b)^k$ est la matrice de dispersion inter-classes pour le mode k calculée à l'aide des vecteurs des paires négatives par:

$$(S^{\text{msida}}_b)^k = \sum_{j=1}^{\prod_{i \neq k} l_i} (S^{\text{msida}}_b)^{k,j}, (S^{\text{msida}}_b)^{k,j} = \sum_{(X_r, X_s) \in D_p} (\mathbf{x}_r^{k,j} - \mathbf{x}_s^{k,j}) (\mathbf{x}_r^{k,j} - \mathbf{x}_s^{k,j})^T \quad (5.7)$$

de même, dans (5.6), $(S^{\text{msida}}_w)^k$ est la matrice de dispersion intra-classe pour le k -mode calculée à l'aide des vecteurs correspondants de la matrice dépliant des paires d'échantillons positifs:

$$(S^{\text{msida}}_w)^k = \sum_{j=1}^{\prod_{i \neq k} l_i} (S^{\text{msida}}_w)^{k,j}, (S^{\text{msida}}_w)^{k,j} = \sum_{(X_o, X_p) \in S_p} (\mathbf{x}_o^{k,j} - \mathbf{x}_p^{k,j}) (\mathbf{x}_r^{k,j} - \mathbf{x}_s^{k,j})^T \quad (5.8)$$

$\mathbf{x}^{k,j}$ désigne le $j^{\text{ième}}$ vecteur de la matrice $\mathbf{X}^{k,j}$ dépliée en mode k d'échantillons de tenseurs \mathbf{X} . Maintenant, la solution est pour un mode est connue, l'optimisation du tenseur entier peut être résolu de manière itérative. Les matrices de projection U_1, U_2, \dots, U_N sont d'abord initialisées à l'identité. À chaque itération $U_1, \dots, U_{k-1}, U_{k+1}, \dots, U_N$ sont supposés connus et U_k est estimé. $\mathbf{Y}_* = \mathbf{X}_* \times_1 U_1 \dots \times_{k-1} U_{k-1} \times_{k+1} U_{k+1} \times_N U_N$, L'équation (5.7) devient:

$$U_k^* = \arg \max_{U_k} \frac{\sum_{D_p=1}^{n_{Dp}} C_j \|Y_r \times_k U_k - Y_s \times_k U_k\|^2}{\sum_{S_p=1}^{n_{Sp}} \|Y_o \times_k U_k - Y_p \times_k U_k\|^2} \quad (5.9)$$

L'équation (5.9) est similaire à (5.6), qui peut être résolu par décomposition en valeurs propres. Le processus itératif de MSIDA s'interrompt lors de la réalisation de l'une des conditions suivantes: soit i) le nombre d'itérations atteint un maximum prédéfini; ou ii) la norme de différence de la projection estimée entre deux itérations successives est inférieure à un seuil, $\|U_k^{itr} - U_k^{itr-1}\| < I_k I_k \varepsilon$, où I_k est la dimension du mode k et U_k^{itr} est la matrice de vecteurs propres en mode k calculée par:

$$(S^{msida})_b^k U_k^{itr} = \Lambda_k (S^{msida})_w^k U_k^{itr} \quad (5.10)$$

où Λ_k est une matrice diagonale dont les éléments diagonaux sont les valeurs propres λ_i^k . Les dernières dimensions inférieures $\hat{I}_1 \times \hat{I}_2 \dots \times \hat{I}_N$ sont définies en fonction du pourcentage d'énergie ($Energy_k$) des valeurs propres à conserver pour chaque mode k:

$$Energy_k = \frac{\sum_{i=1}^{\hat{I}_k} \lambda_i^k}{\sum_{i=1}^{\hat{I}_k} \lambda_i^k} \times 100 \quad (5.11)$$

avec $\lambda_1^k > \lambda_2^k \dots > \lambda_{\hat{I}_k}^k$

La procédure complète pour la méthode proposée MSIDA est fournie dans l'algorithme 5.1.

Algorithme 5.1: Multilinear Side-Information based Discriminant Analysis (MSIDA)

Entrées:

- Tenseur $\tilde{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times N \times M}$ de M échantillons d'apprentissage.
- Les étiquettes faibles pour extraire les paires de tenseur match $S_p = \{(\mathbf{X}_o, \mathbf{X}_p) : l(\mathbf{X}_o) = l(\mathbf{X}_p)\}$ et les paires de tenseur non match $D_p = \{(\mathbf{X}_r, \mathbf{X}_s) : l(\mathbf{X}_r) \neq l(\mathbf{X}_s)\}$ de $\tilde{\mathbf{X}}$.
- Le nombre maximal d'itérations: $I_{tr_{max}}$.
- L'énergie ($Energie_1, Energie_2, \dots, Energie_N$) utilisée pour choisir les dimensions inférieures finales: $I'_1 \times I'_2 \dots I'_N$.

Sorties:

Les matrices de projection des différents modes $U_k = U_k^t \in \mathbb{R}^{I_k \times I'_k}, k = 1, \dots, N$

Algorithme:

1. **Initialisation:** $U_1^0 = I_{I_1}, U_2^0 = I_{I_2}, \dots, U_N^0 = I_{I_N}$
2. **Pour** itr : 1 à $I_{tr_{max}}$
 - (a) **Pour** $k=1$ à N
 - $\mathbf{Y}_i = \mathbf{X}_i \times_1 U_1^{itr-1} \dots \times_{k-1} U_{k-1}^{itr-1} \times_{k+1} U_{k+1}^{itr-1} \dots \times_N U_N^{itr-1}$
 - $Y_i^k \leftarrow \mathbf{Y}_i$
 - $(S^{msida})_b^k = \sum_{j=1}^{\prod_{i \neq k} I_i} (S^{msida})_b^{k,j}, (S^{msida})_b^{k,j} = \sum_{(\mathbf{X}_r, \mathbf{X}_s) \in D_p} (\mathbf{x}_r^{k,j} - \mathbf{x}_s^{k,j}) (\mathbf{x}_r^{k,j} - \mathbf{x}_s^{k,j})^T$
 - $(S^{msida})_w^k = \sum_{j=1}^{\prod_{i \neq k} I_i} (S^{msida})_w^{k,j}, (S^{msida})_w^{k,j} = \sum_{(\mathbf{X}_o, \mathbf{X}_p) \in S_p} (\mathbf{x}_o^{k,j} - \mathbf{x}_p^{k,j}) (\mathbf{x}_r^{k,j} - \mathbf{x}_s^{k,j})^T$
 - Calculer $(S^{msida})_b^k U_k^{itr} = \Lambda_k (S^{msida})_w^k U_k^{itr}$
 - (b) **Si** itr > 2 et $\|U_k^{itr} - U_k^{itr-1}\| < I_k \times I'_k \varepsilon, k = 1, \dots, M$, pause ;
3. Calculer les dimensions inférieures finales I'_k par: $Energie_k = \frac{\sum_{i=1}^{I'_k} \lambda_i^k}{\sum_{i=1}^{I_k} \lambda_i^k} \times 100$, où $(\lambda_1^k > \lambda_2^k \dots > \lambda_{I_k}^k)$.
4. Trier I'_k vecteurs propre $Y_k^{itr} \in \mathbb{R}^{I_k \times I'_k}$ en fonction de λ_i^k par ordre décroissant, $K = 1, \dots, N$

5.4 Vérification des paires de visages basés sur MSIDA

Dans cette section, nous expliquons les détails de l'utilisation du MSIDA proposé pour deux applications de matching de visage différentes : vérification de l'identité et de la parenté à partir de paires d'images de visage. Comme le montre la figure 5.1, le schéma fonctionnel de l'approche proposée comprend trois composants essentiels: l'extraction de caractéristiques, la transformation du sous-espace tenseur et la comparaison.

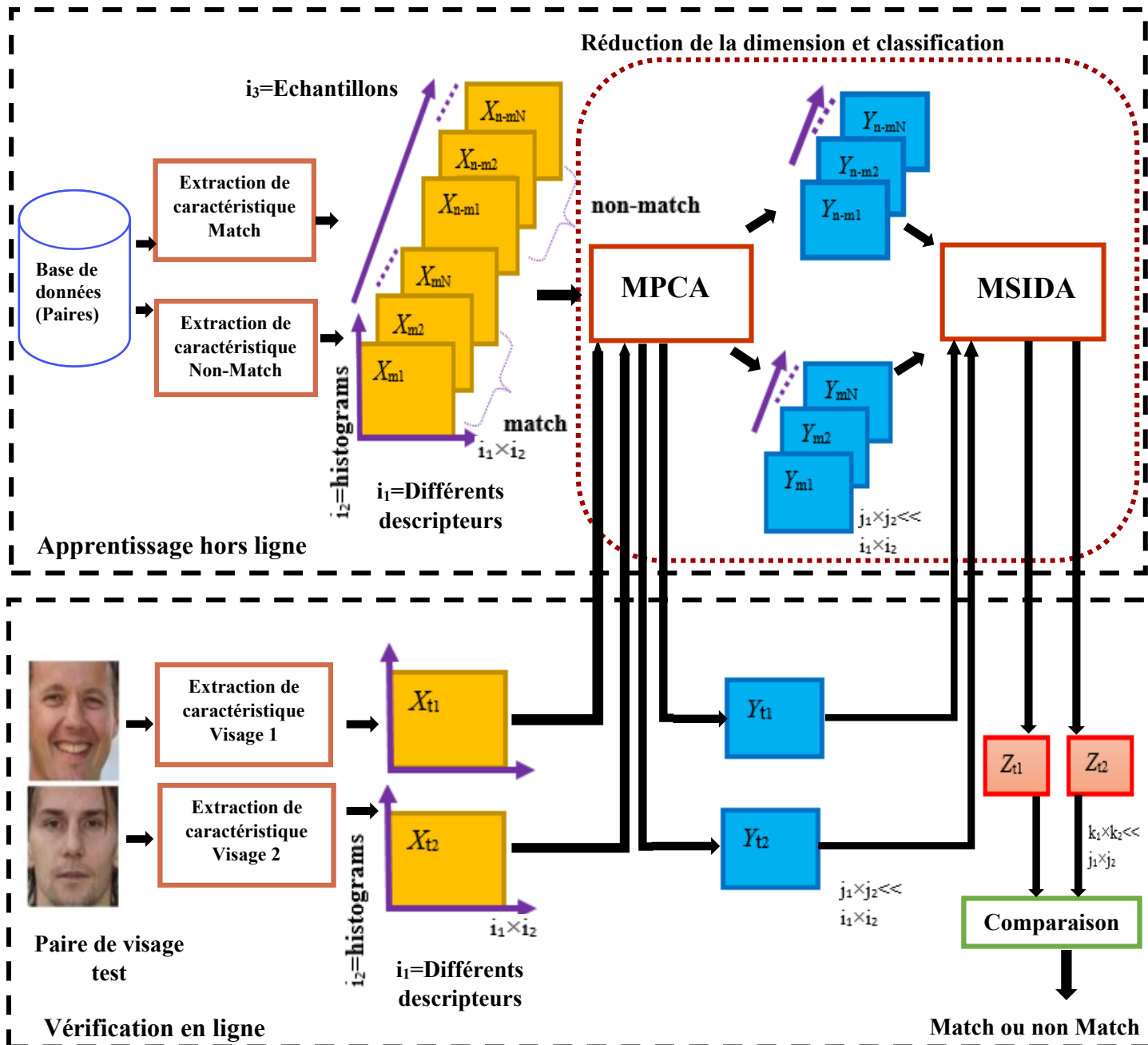


Figure 5.1: Schéma fonctionnel du système de matching des paires de visage proposé.

Chaque image de visage est représentée en utilisant deux descripteurs locaux, MSLPQ et MSBSIF, extraits à différentes échelles, donnant plusieurs vecteurs de caractéristiques par visage. Les vecteurs caractéristiques de tous les visages d'apprentissage sont disposés formant un tenseur de troisième ordre (i_1, i_2, i_3) , où i_1 correspond à un seul vecteur de caractéristiques, i_2 correspond aux différents descripteurs locaux extraits à différentes échelles, et i_3 correspond aux échantillons de visage de la base de données d'apprentissage. Ce tenseur est d'abord projeté avec MPCA dans un sous-espace inférieur $j_1 \times j_2 \times i_3$, où $j_1 \times j_2 \ll i_1 \times i_2$. La raison d'appliquer MPCA avant MSIDA est d'éviter le problème de la petite taille de l'échantillon dans différents modes tensoriels. Ce problème se produit lorsque la dimension des caractéristiques est supérieure au nombre d'échantillons, ce qui conduit à la singularité des matrices de dispersion MSIDA. La réduction de la dimension de chaque mode du tenseur est d'abord effectuée.

Après l'application de MPCA, le tenseur de données d'apprentissage est divisé en deux sous-tenseurs correspondant aux paires matches et aux paires non match, respectivement. La division est effectuée selon le troisième mode i_3 . Le premier tenseur est utilisé pour calculer la matrice de dispersion intra-classe $(S^{msida})_w$ et le deuxième tenseur est utilisé pour calculer la matrice de dispersion inter-classes $(S^{msida})_b$ du MSIDA. Le tenseur de données est projeté avec MSIDA vers un nouveau sous-espace inférieur et discriminant $k_1 \times k_2$, $k_1 \times k_2 \ll j_1 \times j_2$.

Durant la phase de test, chacune des images de visage de la paire est représentée par un tenseur de second ordre formé en empilant les descripteurs locaux de l'image. Par la suite, les deux tenseurs sont projetés par MPCA puis MSIDA. Enfin, la similarité de cosinus entre la paire est calculée et utilisée pour vérifier si la paire est match (appartenant à la même personne / famille) ou non.

Dans ce qui suit, nous expliquons les étapes détaillées du système proposé.

5.4.1 Extraction de caractéristiques

Pour l'extraction de caractéristiques, nous utilisons deux descripteurs locaux: LPQ (Local Phase Quantization) et BSIF (Binarized Statistical Image Feature). LPQ quantifie la phase des basses fréquences tandis que BSIF code les réponses aux filtres pré-appris.

5.4.1.1 Descripteur LPQ (Local Phase Quantization)

Le descripteur LPQ est proposé par Ojansivu et Heikkilä [130] pour l'extraction de caractéristiques des images de textures et de visage 2D. Nous avons présenté en détails ce

descripteur dans la sous-section 4.3.1.1. La figure 5.2 présente une illustration des images LPQ avec différentes échelles dans lesquelles la taille de la fenêtre LPQ prend des valeurs différentes.

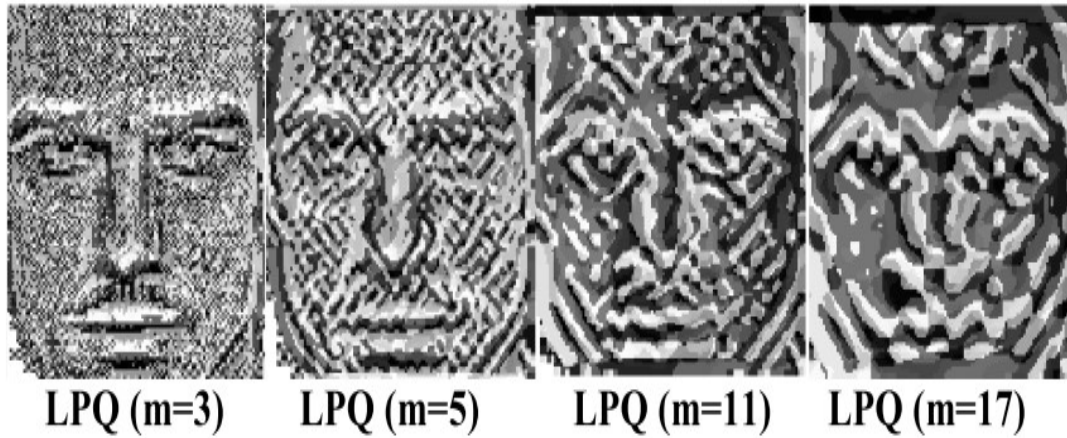


Figure 5.2: Représentation d'image du visage avec le descripteur LPQ sous différentes tailles de fenêtre m .

5.4.1.2 Descripteur BSIF (Binarized Statistical Image Features)

Le descripteur BSIF est proposé récemment par Kannala et al [158] pour la reconnaissance faciale 2D. Dans ce travail, nous essayons d'élargir ce descripteur à la vérification de liens de parenté et la classification de texture. Le descripteur BSIF est inspiré par les deux descripteurs LBP (local binary patterns) et LPQ. Le principe de descripteur BSIF est de calculer un code binaire pour chaque pixel en utilisant un ensemble de filtres linéaires. Cet ensemble de filtres est appris automatiquement à partir d'un ensemble d'apprentissage de 13 images naturelles donné par A. Hyvärinen et al dans [159] (voir figure 5.3) en appliquant l'algorithme ICA (Independent Component Analysis) en maximisant l'indépendance statistique des réponses du filtre. Considérant un patch X de taille $l \times l$ pixels et un filtre linéaire W_i de même taille, la réponse du filtre est donnée par

$$h_i = \sum_{u,v} W_i(u,v) X(u,v) = \mathbf{w}_i^T \mathbf{x} \quad (5.12)$$

où \mathbf{w}_i^T et \mathbf{x} vecteurs contiennent les pixels de W_i et X respectivement. Une chaîne de code binaire b est obtenue en binarisant chaque réponse h_i avec un seuil à zéro comme suit:

$$b_i = \begin{cases} 1 & \text{si } h_i > 0 \\ 0 & \text{autrement} \end{cases} \quad (5.13)$$

La figure 5.4 illustre la représentation BSIF d'une image de profondeur avec différentes taille du filtre l .

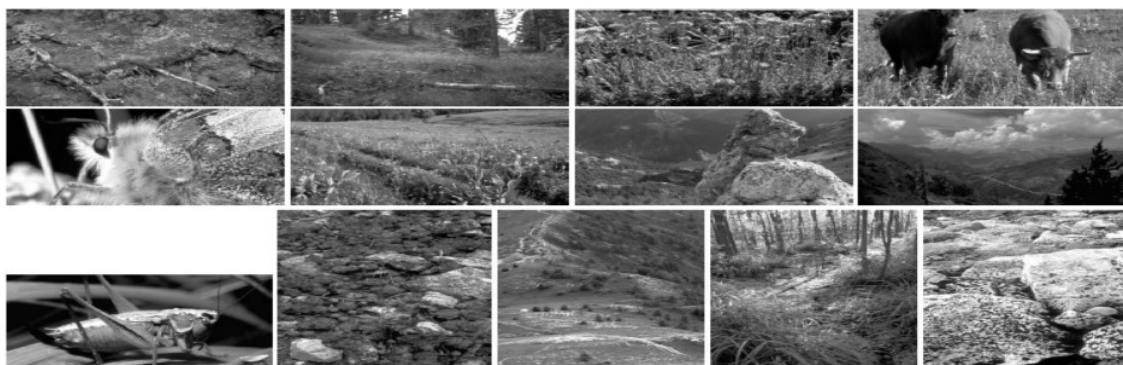


Figure 5.3: Les 13 images naturelles utilisées pour l'apprentissage des filtres dans le descripteur BSIF [159].

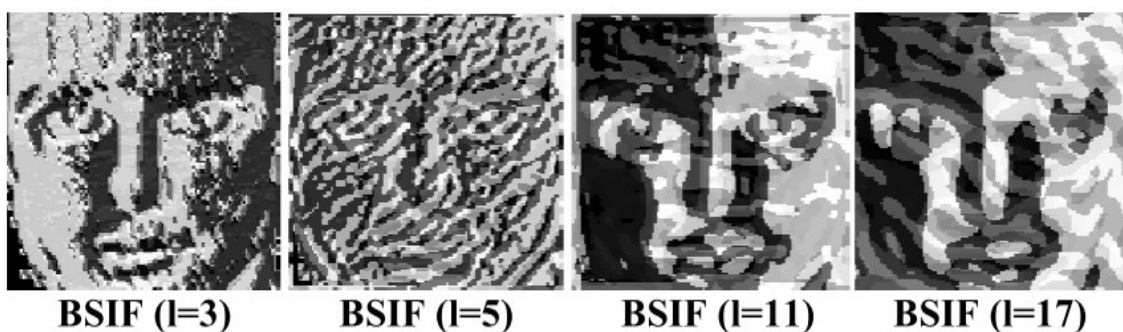


Figure 5.4: Représentation d'image du visage avec le descripteur BSIF sous différentes tailles du filtre l .

Afin de conserver la structure de visage spatiale, l'image de visage est subdivisée en P régions de visage sans chevauchement. Les histogrammes des descripteurs des P régions sont concaténés pour former le vecteur de caractéristiques de visage. Pour enrichir encore la description du visage, nous utilisons la multi-échelle LPQ (MSLPQ) et multi-échelle BSIF (MBSIF). La représentation multi-échelle pour les deux descripteurs LPQ et BSIF peut être obtenue en faisant varier la taille de la fenêtre m et la taille du filtre l , respectivement.

5.4.2 Matching

Pour calculer la similarité entre une paire de visages, nous utilisons la similarité de cosinus [160]. La similarité de cosinus entre deux vecteurs \mathbf{z}_{t_1} et \mathbf{z}_{t_2} est définie comme suit :

$$\cos(\mathbf{z}_{t_1}, \mathbf{z}_{t_2}) = \frac{\mathbf{z}_{t_1}^T \cdot \mathbf{z}_{t_2}}{\|\mathbf{z}_{t_1}\| \cdot \|\mathbf{z}_{t_2}\|} \quad (5.14)$$

où $\|\cdot\|$ est la norme euclidienne. Une valeur élevée du score produit signifie une forte probabilité que la paire de visages appartienne à la même personne / famille.

5.5 Expériences

Dans cette section, nous évaluons l'approche proposée pour deux problèmes de matching de visage, la vérification de l'identité et du lien de parenté dans les environnements non contrôlés. Nous introduisons d'abord les quatre bases de données de visage utilisées pour les expériences. Ensuite, les paramétrages de réglage sont donnés et les résultats sont rapportés et discutés. Enfin, nous comparons nos meilleurs résultats avec l'état de l'art.

5.5.1 Base de données de visage

Quatre bases de données de visage différentes sont prises en compte pour évaluer notre approche. Voici une brève description des données.

5.5.1.1 Base de données LFW (Labeled Faces in the Wild)

LFW [161] est une grande base de données de visage collectée à partir du web. Elle est particulièrement conçue pour étudier le problème de reconnaissance des visages dans des environnements non contrôlés, couvrant les variations du monde réel en termes de pose, d'illumination, d'expressions, de résolution, de flou, d'occlusion, etc. Cette base de données comprend 13,233 images de 5,749 personnes différentes. La base de données est divisée en deux vues: la vue 1 utilisée pour la sélection du modèle et la vue 2 utilisée pour l'évaluation des performances. LFW définit trois protocoles d'évaluation: restreint, non restreint et non supervisé. Dans nos expériences, nous évaluons l'approche proposée sur la vue 2 en utilisant des images de visage alignées selon le protocole avec restriction, où aucune donnée d'apprentissage extérieure ne peut être impliquée. La base de données est divisée en 10 sous-ensembles disjoints pour la validation croisée, 9 sous-ensembles étant utilisés pour l'apprentissage et le sous-ensemble restant pour le test. Chaque sous-ensemble contient 300 paires match et 300 paires non match. La performance est indiquée sous forme de précision moyenne et de courbe ROC sur 10 validations croisées. Des exemples d'images de visage de cette base de données sont présentés à la figure 5.5.



Figure 5.5: Exemples d'images de la base de données LFW. Gauche : paires positives et droite : paires négatives.

5.5.1.2 Base de données Cornell KinFace

Cornell KinFace [151] comprend 150 paires d'images de visages de personnes ayant une relation de parenté. La base de données a été collectée à partir d'Internet en considérant quatre relations familiales. La distribution des relations familiales dans l'ensemble de données est de 40% père fils (P-Fs), 22% père-fille (P-F), 13%, mère-fils (M-Fs) et 25% mère-fille (M-F). Chaque paire est composée d'une image de visage parent et d'une image de visage enfant. Dans nos expériences, nous évaluons l'approche proposée sur le même protocole que celui défini par Yan et al [162], où seulement 143 paires sont utilisées et la classification est réalisée avec 5 validations croisées pour la vérification de la parenté. Les paires négatives sont générées aléatoirement en associant chaque image parent à une image d'un autre enfant. Des exemples d'images de visage de cette base de données sont illustrés à la figure 5.6.



Figure 5.6: Exemples d'images de la base de données Cornell KinFace. Gauche: paires positives et droite: paires négatives.

5.5.1.3 Base de données UB KinFace

UB KinFace [163] comprend 600 images de 400 personnes, collectée à partir du web. Ces images sont séparées en 200 familles. Chaque famille est composée de trois images,

correspondant respectivement à l'enfant, au jeune parent et au parent âgé. Cela forme deux sous-ensembles de paires: ensemble 1 de 200 enfants et 200 jeunes parents et ensemble 2 de 200 enfants et 200 parents âgés. Dans nos expériences, nous évaluons l'approche proposée sur le même protocole que celui défini par Yan et al [162], en utilisant 5 validations croisées. Des exemples d'images de visage de cette base de données sont présentés à la figure 5.7.



Figure 5.7: Exemples d'images de la base de données UB KinFace. Gauche : paires positives et droite : paires négatives.

5.5.1.4 Base de données TSKinFace

TSKinFace [164] est une base de données récemment publiée, et elle est parmi les plus grandes bases de données de parenté disponibles. Au total, la base de données compte 4060 images de visage. Il comprend deux ensembles de trois relations de famille: Père-Mère-Fils (PM-Fs) et Père-Mère-Fille (PM-F). Il y a 513 PM-F et 502 PM-F groupes distincts de trois sujets. Dans notre cas, nous avons restructuré la base de données en scindant l'ensemble Père-Mère-Fils en deux paires. Père-Fils et Mère-Fils et l'ensemble Père-Mère-Fille en deux paires Père-Fille et Mère-Fille. En suivant le même protocole de vérification de la parenté que celui défini par Qin et al [164], nous avons utilisé 5 validations croisées. Chaque ensemble comprend à peu près le même nombre de paires de visages. Des exemples d'images de visage de cette base de données sont présentés à la figure 5.8.



Figure 5.8: Exemples d'images de la base de données TSKinFace. Gauche : paires positives et droite : paires négatives.

5.5.2 Paramètres de réglage

Les paramètres donnant les résultats rapportés de nos expériences sont fournis dans ce qui suit. Les images de visage sont recadrées et redimensionnées à 130×90 pixels, 126×115 pixels, 96×80 pixels et 64×64 pixels pour les bases de données LFW, Cornell, UB KinFace et TSKinFace, respectivement. Après avoir extrait MSLPQ et MSBSIF de chaque image de visage, l'image résultante est subdivisée en 20 blocs sans chevauchement. Pour extraire les deux descripteurs à différentes échelles, nous modifions la taille de la fenêtre (m) de LPQ et la taille du filtre (l) de BSIF. Les mêmes valeurs sont considérées pour les deux paramètres m et $l = \{3, 5, 7, 9, 11, 13, 15, 17\}$. Pour conserver le vecteur de caractéristique de taille raisonnable tout en maximisant la description de l'image, le nombre d'échelles combinées simultanément varie de 2 à 4 pour les deux descripteurs. Ces paramètres sont adoptés, car leurs résultats empiriques s'avèrent être les meilleurs.

Les données d'apprentissage sont utilisées pour estimer les matrices de projection de sous-espace. Le sous-espace MSIDA est défini sur 96% de l'énergie des valeurs propres (comme l'indique l'équation (5.8), les valeurs propres qui permettent d'économiser jusqu'à 96% des informations sont conservées). Le nombre d'itérations, pour les algorithmes MPCA et MSIDA, est réglé de manière empirique et la meilleure valeur est $Itr_{\max} = 2$.

5.5.3 Résultats et discussion

Cette sous-section présente les résultats de l'évaluation et la discussion. En plus de la représentation tenseur de troisième ordre, toutes les expériences sont également menées pour l'approche linéaire originale, SILD, qui sert de base à l'évaluation de la méthode proposée, MSIDA. De plus, nous examinons individuellement les performances des descripteurs locaux, MSLPQ et MSBSIF, ainsi que leur fusion. Dans le cas linéaire, la fusion de niveaux de caractéristiques est réalisée en concaténant des vecteurs d'échelles différentes pour chaque descripteur de visage. Pour le cas multilinéaire proposé, la fusion est effectuée sur le tenseur, où des vecteurs d'échelles différentes de deux descripteurs, LPQ et BSIF, sont empilés dans le deuxième mode du tenseur.

Pour atténuer le problème de la taille réduite de l'échantillon, les dimensions de caractéristiques sont d'abord réduites comme suit. Dans le cas d'une représentation linéaire (tenseur du second ordre), nous appliquons PCA avant SILD. Dans le cas d'une représentation multilinéaire (tenseur de troisième ordre), nous appliquons MPCA avant MSIDA.

Tableau 5.1: Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données LFW.

Caractéristique	Précision moyenne (%) \pm Erreur standard (%)	
	PCA+SILD	MPCA+MSIDA
MSLPQ (m=3+5)	89.60 \pm 1.21	93.07 \pm 1.06
MSLPQ (m=5+9+13)	90.43 \pm 1.25	94.33 \pm 0.94
MSLPQ (m=13+15+17)	90.33 \pm 1.21	94.02 \pm 1.03
MSLPQ (m=5+7+9+11)	90.03 \pm 1.28	93.57 \pm 1.04
MBSIF (m=3+5)	90.27 \pm 1.39	92.27 \pm 1.34
MBSIF (m=5+9+13)	89.00 \pm 1.29	94.05 \pm 0.95
MBSIF (m=13+15+17)	90.47 \pm 1.21	94.37 \pm 0.85
MBSIF (m=5+7+9+11)	89.60 \pm 1.11	93.03 \pm 1.14
MSLPQ (m=5+9+13) + MSBSIF (l=13+15+17)	91.30 \pm 1.26	94.40 \pm 0.89

Tableau 5.2: Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données Cornell KinFace.

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

Tableau 5.3: Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données UB KinFace.

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

Tableau 5.4: Précision de la vérification de MSIDA et SILD en utilisant différentes échelles MSLPQ et MSBSIF et leur fusion sur la base de données TSKinFace.

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

Les résultats, en termes de précision de la vérification, sont présentés dans le tableau 5.1 pour la vérification de visage dans la base de données LFW et dans les tableaux 5.2, 5.3 et 5.4 pour la vérification de la parenté dans les ensembles de données Cornell KinFace, UB KinFace et TSKinFace, respectivement.

Du point de vue des descripteurs et au niveau individuels, nous pouvons remarquer que, dans la plupart des cas, MSBSIF obtient de meilleurs résultats que MSLPQ. Nous pouvons également observer que les meilleurs résultats sont obtenus pour MSBSIF avec une combinaison de filtres

de grande taille (13 + 15 + 17), tandis que la combinaison de fenêtres de taille moyenne (5 + 9 + 13) des descripteurs MSLPQ donne les meilleurs résultats. En outre, la fusion des caractéristiques des trois filtres précédents des deux descripteurs présente de manière permanente de meilleurs résultats dans les quatre tableaux. Par conséquent, nous sélectionnons ces deux paramètres multi-échelles (13 + 15 + 17 pour MSBSIF et 5 + 9 + 13 pour MSLPQ) lors de la fusion des deux descripteurs. Cette dernière fusion donne dans tous les cas de meilleurs résultats que les descripteurs individuels, démontrant ainsi la complémentarité de MSBSIF et de MSLPQ pour la description du visage.

Considérant l'influence des approches de sous-espace, pour chaque cas particulier des tableaux 5.1, 5.2, 5.3 et 5.4 (quels que soient les caractéristiques, les échelles, les bases de données et le problème de matching de visage), l'approche multilinéaire proposée MSIDA donne de meilleurs résultats que son contrepartie linéaire SILD. Une analyse plus approfondie des résultats révèle que MSIDA est capable d'extraire meilleure caractéristique discriminante que SILD.

Dans le cas linéaire MSBSIF fonctionne toujours mieux que MSLPQ, avec une marge perceptible, la différence de performance entre les deux descripteurs est réduite dans le cas multilinéaire.

Les résultats montrent que la méthode multilinéaire faiblement supervisée MSIDA tire pleinement parti de la structure de tenseur d'ordre supérieur. La malédiction de dimensionnalité est intuitivement évitée (problème de la taille réduite de l'échantillon). En effet, dans le cas de la méthode SILD, les vecteurs de caractéristiques sont simplement concaténés en ignorant la structure de données naturelle, ces caractéristiques sont empilées dans un mode tenseur permettant l'extraction d'informations maximales. De plus, MSIDA extrait mieux les informations discriminantes présentes dans chaque mode par rapport à la méthode SILD, qui extrait les informations discriminantes par une seule projection.

5.5.4 Effet de la quantité de paires d'apprentissage

Nous étudions l'effet de la quantité de paires d'entraînement pour les bases de données LFW, Cornell KinFace, UB KinFace et TSKinFace. Différentes quantités de paires d'images d'apprentissage sélectionnées aléatoirement sont utilisées dans cette expérience. Plus précisément, 600, 1200, 1800, 2400, 3000, 3600, 4200, 4800 et 5400 paires sont utilisées pour la base de données LFW. Pour les bases de données Cornell et UB KinFace, nous avons utilisé

30, 60, 90, 120, 150, 180 et 210 paires. Pour TSKinFace, nous avons utilisé 80, 160, 240, 320, 400, 480, 560, 640, 720 et 800 paires. La figure 5.9 montre la précision moyenne de la méthode proposée MSIDA et de son contrepartie SILD par rapport le nombre de paires d'apprentissage. Nous pouvons clairement observer que les performances des deux méthodes MSIDA et SILD deviennent meilleures lorsque le nombre de paires d'apprentissage augmente. De plus, notre méthode MSIDA surpasse SILD sur toutes les bases de données pour différentes quantités de données d'apprentissage. Les résultats expérimentaux indiquent que la quantité de paires d'apprentissage a une influence significative sur la précision finale. Ceci est démontré par les améliorations de précision entre deux configurations successives de l'expérience.

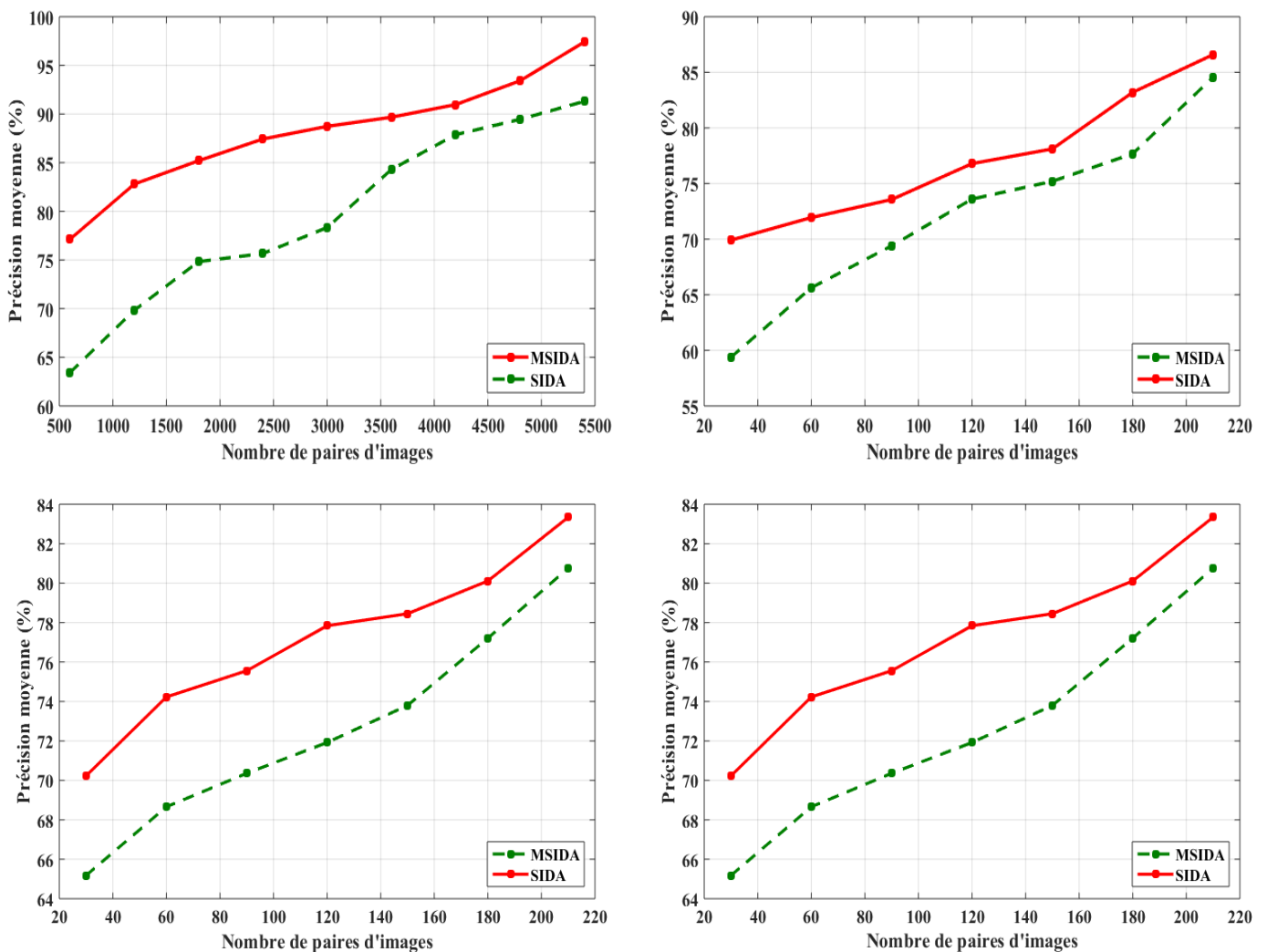


Figure 5.9: Précision moyenne de SILD et MSIDA avec un nombre différent de paires sur les bases de données (a) LFW (b) Cornell KinFace (c) UB KinFace (d) TSKinFace, respectivement.

5.5.5 MSIDA faiblement supervisé contre LDA et MDA supervisé

Dans cette expérience, nous comparons le SILD de base et la méthode proposée MSIDA sous le protocole restreint avec LDA et MDA sous le protocole non restreint sur la base de données LFW. La précision de la vérification des différentes méthodes MSIDA, SILD, MDA et LDA est indiquée dans le tableau 5.5. Leurs courbes ROC sont illustrés dans la figure 5.10. Nous pouvons observer que MSIDA et SILD offrent des performances comparables à celles de leurs contreparties supervisés, MDA et LDA. De plus, la marge d'amélioration de la MDA en présence d'étiquettes d'identité est plus petite par rapport au paramètre d'information faible avec MSIDA (seule une petite différence inférieure à 0,67%). Par conséquent, nous pouvons confirmer que MSIDA avec faible d'informations atteint des performances comparables avec l'utilisation de l'étiquette de classe complète dans le cas supervisé avec MDA. De plus, avec des informations faiblement étiquetées, notre méthode MSIDA est meilleure que LDA avec des informations étiquetées complètes. Cet avantage présente le MSIDA proposé comme un modèle efficace pour traiter le problème de faible information.

Tableau 5.5: Précision de la vérification (%) des différentes méthodes selon les protocoles restreint et non restreint sur la base de données LFW.

Paramètre	Faiblement supervisé (sans étiquette de classe complète)		Supervisé (avec étiquette de classe complète)	
	Restreint		Non restreint	
Méthode	MSIDA	SILD	MDA	LDA
Précision moyenne (%) ± Erreur standard (%)	94.40 ± 0.89	91.30 ± 1.26	95.07 ± 0.71	91.93 ± 0.82

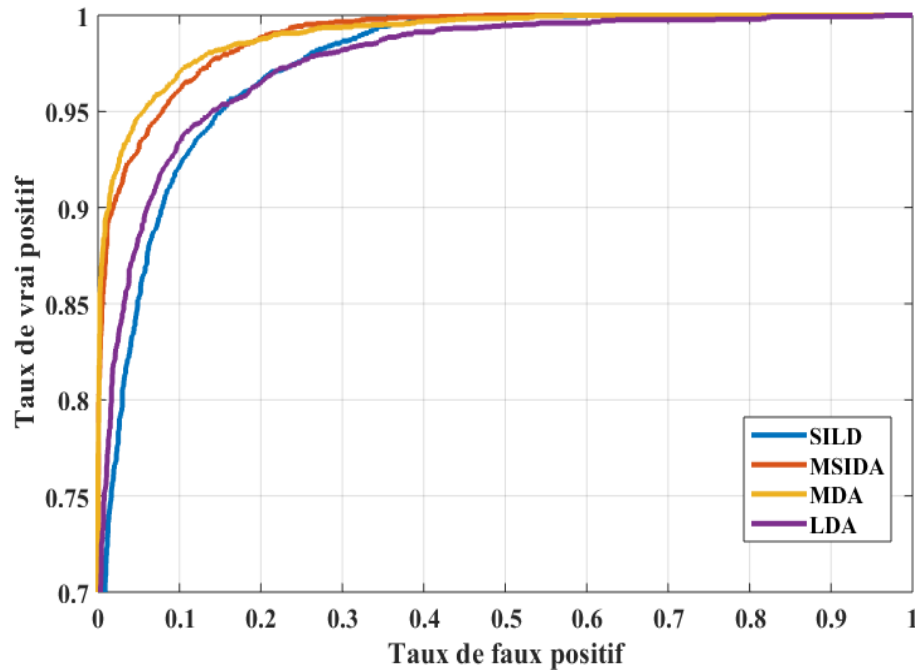


Figure 5.10: Courbes ROC des différentes méthodes selon les protocoles restreint et non restreint sur la base de données LFW.

5.5.6 Temps de complexité

Les expériences ont été réalisées sous MATLAB version 2015a sur un PC doté d'un processeur Intel Xeon (R) à 3,19 GHz et de 12 Go de RAM. L'estimation des matrices de projection est implémentée dans la phase d'apprentissage hors ligne. La complexité en termes de temps de calcul (TC) pour une matching de paire de visages (pendant la phase en ligne) avec SILD et du MSIDA proposé est présentée dans le tableau 5.6. Depuis les mêmes caractéristiques MSLPQ + MSBSIF) sont utilisés pour décrire l'image du visage à la fois pour SILD et MSIDA, le même TC est requis pour les deux méthodes à cette étape. Pour la réduction de la dimensionnalité et la comparaison, on peut voir que notre méthode MSIDA réalise un bon gain en TC par rapport à la méthode de base dans tous les bases de données. Le TC est réduit grâce à l'utilisation de la représentation tensorielle des données de visage. En plus de la précision, cette expérience démontre l'efficacité de notre approche en termes de complexité de calcul. En effet, le MSIDA ne dépasse pas un total TC de 0,1259 s, ce qui en fait un candidat compétent pour l'utilisation dans des applications de vérification du visage dans le monde réel.

Tableau 5.6: Temps de complexité en secondes de SILD et MSIDA sur les quatre bases de données.

Base de données	Extraction de caractéristiques	Réduction de la dimensionnalité et la comparaison		Temps total d'exécution	
		SILD	MSIDA	SILD	MSIDA
LFW	0.1085	0.0217	0.0174	0.1302	0.1259
Cornell KinFace	0.1249	0.0045	0.0032	0.1294	0.1281
UB KinFace	0.0978	0.0050	0.0036	0.1028	0.1014
TSKinFace	0.0785	0.0041	0.0029	0.0826	0.0814

5.5.7 Comparaison avec l'état de l'art

5.5.7.1 Vérification de visage dans les environnements non contrôlés

Notre critère de comparaison est la précision obtenue par les travaux récents non seulement du côté de mécanismes similaires. La comparaison de notre approche en termes d'apprentissage métrique et de type de caractéristiques est fournie dans les tableaux 5.1, 5.2, 5.3, 5.4 et 5.5. Dans cette sous-section, le tableau 5.7 compare nos résultats avec l'état de l'art, y compris les travaux les plus récents sur la base de données LFW. De plus, la figure 5.11 illustre les courbes ROC correspondantes. La meilleure précision de vérification de notre travail est de 94.40%. Ce résultat est classé deuxième des différentes méthodes de l'état de l'art. Les travaux récents MRF-Fusion-CSKDA [38] offrent la meilleure précision de vérification actuellement sur la base de données LFW. MRF-Fusion-CSKDA est basé sur la fusion de trois descripteurs multi-échelles MSLBP, MSLPQ et MSBSIF avec méthode de noyau. En revanche, nous n'utilisons que deux descripteurs, MSLPQ et MSBSIF en utilisant l'approche multilinéaire proposée MSIDA se classant au deuxième rang dans la littérature. Par conséquent, notre système est moins complexe, efficace que la méthode de classement supérieur.

Tableau 5.7: Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données LFW selon le protocole restreint.

Méthode	Précision moyenne (%) ± Erreur standard (%)
Eigenfaces, original [121]	60.02 ± 0.79
Nowak2, funneled [165]	73.93 ± 0.49
Hybrid descriptor-based, funneled [166]	78.47 ± 0.51
Pixels/MKL, funneled [167]	68.22 ± 0.41
Fisher vector faces [168]	87.47 ± 1.49
Eigen-PEP [169]	88.97 ± 1.32
Class-specific kernel discriminant analysis-MRF-MSLPQ [170]	92.42 ± 1.03
Class-specific kernel discriminant analysis-MRF-MSBSIF [170]	93.63 ± 1.27
Class-specific kernel discriminant analysis MRF-Fusion [170]	95.89 ± 1.94
Hierarchical-PEP [171]	91.10 ± 1.47
Side-Information based Linear Discriminant Analysis-MSBSIF [133]	90.37 ± 1.19
Robust Statistical Frontalization [172]	88.81 ± 0.78
Discriminative Deep Multi-Metric Learning [173]	93.28 ± 0.39
MSIDA-MLPQ	94.33 ± 0.94
MSIDA-MBSIF	94.37 ± 0.85
MSIDA-fusion (MLPQ+MBSIF)	94.40 ± 0.89

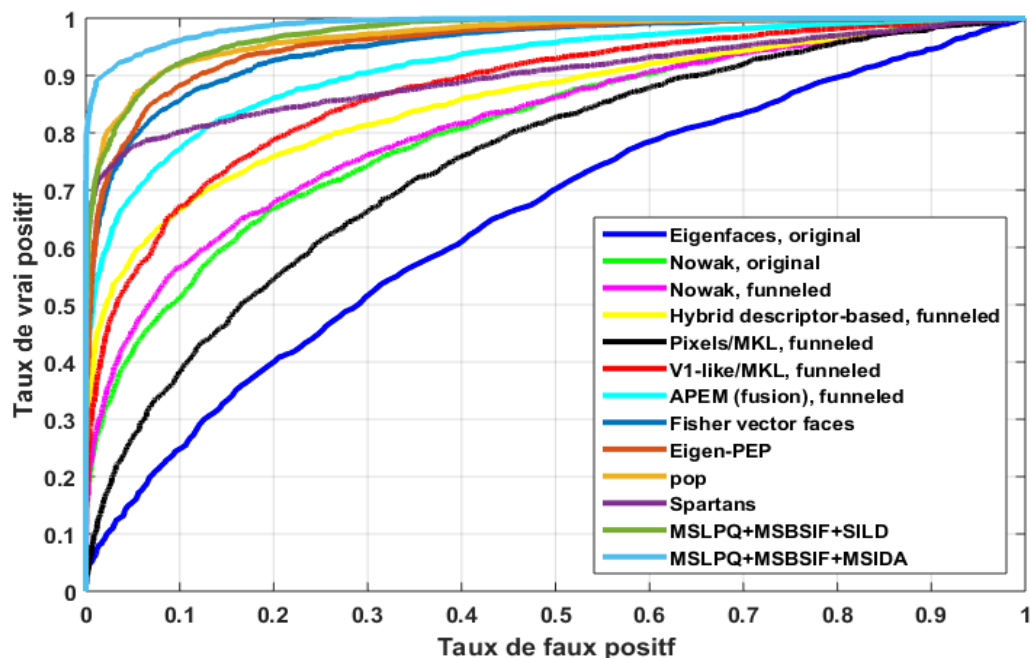


Figure 5.11: Courbe ROC de MSIDA et l'état de l'art sur la base de données LFW sous le protocole le restreint.

5.5.7.2 Vérification de parenté dans les environnements non contrôlés

Les tableaux 5.8, 5.9 et 5.10 comparent l'approche proposée avec l'état de de l'art dans les bases de données Cornell KinFace, UB KinFace et TSKinFace, respectivement. Nous notons que les meilleures précisions de vérification sont obtenues par notre approche de 86,87% sur Cornell KinFace, 83,34% sur UB KinFace et 85,18% sur TSKinFace. Comme le montrent les trois tableaux de comparaison, notre approche surpasse les autres méthodes de l'état de de l'art sur les trois bases de données de la parenté. De plus, nous pouvons clairement voir que MSIDA améliore les performances de la parenté avec une marge significative (par exemple, plus de 13% sur Cornell KinFace et plus de 11% sur UB KinFace). Puisque notre approche est similaire à plusieurs approches, la puissance de notre approche est obtenue grâce à la stratégie multilinéaire basée sur la méthode MSIDA proposé utilisant une représentation tenseur d'ordre élevé qui exploite des informations plus discriminantes.

Tableau 5.8: Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données Cornell KinFace.

Méthode	Précision moyenne (%)
Pictorial structure model-HOG [151]	70.67
Discriminative multimetric learning-fusion (LBP, SIFT, SPLE) [162]	73.50
Multiview Neighborhood Repulsed Metric Learning-fusion (LBP, SIFT, TPLBP and LE) [152]	71.60
Prototype discriminative feature learning-fusion (LBP, SIFT) [174]	71.90
Multiple Kernel Similarity Metric-LPQ [175]	81.70
MSIDA-MLPQ	86.11
MSIDA-MBSIF	86.45
MSIDA-fusion (MLPQ+MBSIF)	86.87

Tableau 5.9: Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données UB KinFace.

Méthode	Précision moyenne (%)
Transfer subspace learning-Gabor [176]	56.50
Self-similarity representation of Weber faces-DOG [177]	53.90
Discriminative multimetric learning-fusion (LBP, SIFT, SPLE) [162]	72.25
Multiview Neighborhood Repulsed Metric Learning-fusion (LBP, SIFT, TPLBP and LE) [152]	67.05
Prototype discriminative feature learning-fusion (LBP, SIFT) [174]	67.30
Random Subset Feature Selection-fusion (LPQ, LBP, SPLE) [178]	76.65
MSIDA-MLPQ	82.97
MSIDA-MBSIF	83.14
MSIDA-fusion (MLPQ+MBSIF)	83.34

Tableau 5.10: Comparaison la précision de vérification de MSIDA avec l'état de l'art sur base de données TSKinFace.

Méthode	Précision moyenne (%)
Relative Symmetric Bilinear Model-SIFT [164]	81.85
Binarized Statistical Image Feature [179]	74.46
Discriminative Deep Multi-Metric Learning-LPQ [173]	79.92
Multiple Kernel Similarity Metric-LPQ [175]	81.89
MSIDA-MLPQ	83.98
MSIDA-MBSIF	83.24
MSIDA-fusion (MLPQ+MBSIF)	85.18

5.6 Conclusion

Dans ce chapitre, nous avons présenté une approche efficace pour le matching des paires de visages dans les environnements non contrôlés. L'approche proposée est basée sur la représentation au tenseur d'ordre élevé des images de visage. Le tenseur du visage est construit à l'aide d'histogrammes de deux descripteurs locaux, MBSIF et MLPQ, extraits à plusieurs échelles. Pour projeter le tenseur du visage dans un sous-espace de faible dimension et discriminant, en tirant parti des données faiblement étiquetées disponibles, nous avons étendu le SILD linéaire à MSIDA, qui fonctionne sur des données multilinéaires. MSIDA trouve des projections multilinéaires du tenseur, où la séparation entre les classes de données est améliorée en exploitant les informations faibles disponibles. L'évaluation expérimentale de notre approche de vérification de l'identité et de la parenté démontre que la méthode proposée est

plus efficace, en termes de précision et de temps de calcul, que sa forme linéaire initiale. De plus, notre approche surpasse l'état de l'art sur trois bases de données de parenté (Cornell KinFace, UB KinFace et TSKinFace) et se classe deuxième sur la base de données LFW.

Chapitre 6

Conclusion générale et Perspectives

6.1 Conclusion générale	123
6.2 Perspectives	125

6.1 Conclusion générale

La reconnaissance de visage est un domaine de recherche très actif en vision par ordinateur et en biométrie en particulier au cours des dernières décennies. La recherche d'un système de reconnaissance faciale automatique robuste est un problème épineux en raison de différentes variations des visages humains, du nombre insuffisant d'échantillons de données et de la complexité de la distribution des motifs. L'apprentissage de sous-espace, ou la réduction de dimensionnalité, tente de projeter des données de grande dimension dans un espace de faible dimension afin d'améliorer la tâche de la reconnaissance automatique. Cependant, les algorithmes traditionnels d'apprentissage de sous-espaces, tels que PCA, LPP et LDA...etc., sont des méthodes linéaires. Ils doivent transformer le visage, qui possède des représentations tensorielles naturelles, en vecteurs de très haute dimension avant l'apprentissage. Cette transformation casse la structure naturelle et les corrélations dans les données multidimensionnelles originales. Cela entraîne également une forte demande en calcul et en mémoire, ainsi que la nécessité d'estimer un très grand nombre de paramètres. En revanche, l'apprentissage de sous-espace multilinéaire traite les objets tensoriels dans leur forme originale et permet de faire l'apprentissage sur des représentations plus utiles et compactes pour une meilleure reconnaissance. Par conséquent, cette thèse a été consacrée au problème de la reconnaissance automatique de visages par l'analyse multidimensionnelle. Elle a contribué au développement d'algorithmes d'apprentissage sous-espaces multilinéaires.

Dans le chapitre 1, nous avons donné une introduction générale des contextes, les motivations, les objectifs et les contributions de cette thèse.

Dans le chapitre 2, nous avons rappelé les principales définitions et propriétés relatives aux tenseurs ainsi que les différents types des projections tensorielles : projection vecteur en vecteur, projection tenseur en vecteur et projection tenseur en tenseur. D'autre part, nous avons présenté le problème d'apprentissage de sous-espace dans les deux cas : linéaire et multilinéaire

Dans le chapitre 3, nous avons passé en revue la littérature des algorithmes de reconnaissance faciale par l'analyse multidimensionnelle basée sur les approches de sous-espaces multilinéaires. Le but de ce chapitre est d'évaluer les capacités et les lacunes de ces techniques de sous-espace multilinéaire dans la reconnaissance de visage, afin d'identifier les problèmes et les orientations pour la recherche future dans ce domaine, et faire un choix judicieux pour la conception du système d'analyse d'images de visage.

Dans le chapitre 4, nous avons présenté un nouvel algorithme supervisé de réduction de la dimensionnalité, pour la reconnaissance faciale multimodale 2D+3D appelé MEFDA (Multilinear Enhanced Fisher Discriminant Analysis). MEFDA est une extension naturelle de la méthode classique linéaire EFM en forme multilinéaire, Alors que l'EFM effectue la réduction de la dimensionnalité dans un seul mode de l'image, la méthode proposée MEFDA cherche d'obtenir les matrices de projection optimales dans chaque mode de données tensorielles. MEFDA respecte la structure d'origine et la géométrie naturelle des données de visage. Par conséquent, la corrélation naturelle des données de visage multidimensionnelles est bien exploitée. De ce fait, MEFDA présente de nombreux avantages importants par rapport à la méthode EFM. Tout d'abord, il cherche à prendre en compte le bon pouvoir de la discrimination de l'EFM et permet d'exploiter les relations cachées entre les vecteurs de lignes et de colonnes de l'image, ce qui améliore la séparation entre les classes de données. Deuxièmement, il améliore les performances de reconnaissance de l'EFM. En outre, l'extraction des caractéristiques de l'image est plus simple grâce à la représentation tensorielle au cours de l'apprentissage, de sorte que moins de temps de calcul est requis. Des expériences approfondies ont été réalisées pour évaluer les performances de MEFDA en utilisant les trois bases de données du visage humain les plus connus FRGC v2.0, CASIA 3D et Bosphorus. Il est bien prouvé que MEFDA surpasse plusieurs méthodes de la réduction de la dimensionnalité telle que PCA, LDA, EFM, MPCA et MDA sur les différentes bases de données en ce qui concerne le taux de reconnaissance et le temps de calcul.

Dans le chapitre 5, nous avons étendu naturellement la méthode de réduction de la dimensionnalité SILD à sa forme multilinéaire en développant une nouvelle approche semi supervisé de la vérification de visage et la parenté appelée MSIDA (Multilinear Side-Information based Discriminant Analysis), qui agit directement sur les données tensorielles naturelles plutôt que sur des vecteurs ou des matrices. Il améliore les performances de la vérification de visage et la parenté rapport au SILD qui ne fonctionne que sur les vecteurs 1D. La méthode proposée MSIDA permet d'obtenir plusieurs sous-espaces de projection interreliés en déployant les données multidimensionnelles sur différentes modes de tenseurs. Le nombre de sous-espaces dérivés par MSIDA dépend de l'ordre de l'espace de données tensoriel. En général, la méthode proposée MSIDA présente les caractéristiques suivantes :

- Il résout le problème de la projection et de la classification des sous-espaces multilinéaires faiblement supervisés. En utilisant uniquement les informations d'étiquette faibles, MSIDA projette le tenseur des visages d'entrée dans un nouveau

sous-espace dans lequel la marge entre des échantillons appartenant à différentes classes est agrandi alors que la marge des échantillons appartenant aux mêmes classes est réduite.

- Il dérive les sous-espaces de manière itérative en dépliant le tenseur dans différentes modes pour l'extraction plus des informations discriminantes et la classification des données.
- Il préserve la structure locale des échantillons de données dans l'espace des caractéristiques, à l'aide des représentations tensorielles d'ordre supérieur, afin de révéler les structures essentielles pour l'analyse de données, sans nécessiter de vectorisation des données.
- contrairement à sa contrepartie traditionnelle SILD, la méthode proposée évite efficacement les problèmes de petite taille d'échantillon.

La performance de la méthode MSIDA a été comparée à celle de méthodes d'apprentissage de sous-espaces telles que SILD, LDA et MDA. La méthode proposée a été validée par plusieurs types d'analyses en utilisant quatre bases de données de visage et la parenté : LFW, Cornell KinFace, UB KinFace et TSKinFace. Les résultats de l'étude expérimentale montrent que MSIDA est une méthode de reconnaissance robuste et efficace pour les objets tensoriels, car il surpasse toutes les approches standards.

6.2 Perspectives

Bien que de nombreux problèmes de reconnaissance faciale par l'analyse multidimensionnelle aient été traités dans cette thèse, le domaine de l'apprentissage de sous-espaces multilinéaire présente encore de nombreux problèmes non résolus qui doivent encore être pris en compte. Par conséquent, cette section présente certains sujets pour des recherches futures dans le but de concevoir des solutions d'apprentissage de sous-espaces multilinéaires, ainsi de développer de nouvelles applications dans lesquelles les méthodes proposées peuvent être appliquées.

Développement d'algorithmes d'apprentissage des sous-espaces multilinéaires

- Le traitement systématique de l'apprentissage de sous-espace multilinéaire présenté au chapitre 2 bénéficiera au développement de nouveaux algorithmes d'apprentissage multilinéaire, notamment en étendant les idées riches pour proposer d'autres algorithmes en utilisant des fonctions noyaux. Dans les travaux futurs, une stratégie plus

complexe (non linéaire) peut être obtenue en développant des extensions multidimensionnelles non linéaires pour MEFDA et MSIDA.

- La solution dans les deux approches MEFDA et MSIDA est basée sur la projection tenseur en tenseur. Il serait intéressant de dériver une solution multilinéaire pour ces méthodes proposées basée sur la projection tenseur en vecteur (V-MEFDA/V-MSIDA). Par conséquent, cette stratégie peut être plus appropriée pour les tâches de reconnaissance où un petit nombre de caractéristiques est nécessaire.
- La fusion des caractéristiques est devenue l'une des méthodes les plus efficaces pour améliorer les performances de reconnaissance. La combinaison aux chapitres 4 et 5 de MEFDA et MSIDA avec les descripteurs locaux MLPQ et MBSIF a montré des résultats prometteurs pour la vérification faciale et la parenté. Il serait donc intéressant d'étudier l'intégration des vecteurs de caractéristiques extraits par MEFDA et MSIDA avec d'autres caractéristiques locaux afin d'extraire les informations les plus discriminantes pour une meilleure séparation.
- Enfin, dans l'apprentissage des sous-espaces multilinéaires, il reste encore de nombreux problèmes non résolus, tels que l'initialisation optimale, l'ordre de projection optimal et le critère d'arrêt optimal. Cette thèse a fait quelques tentatives pour résoudre certains de ces problèmes dans les deux méthodes proposées MEFDA et MSIDA. Cependant, il sera bénéfique que des recherches plus approfondies conduisent à une compréhension plus approfondie de ces questions.

Explorer d'autres applications des algorithmes d'apprentissage de sous-espaces multilinéaires

En plus de la représentation tensorielle du visage étudiée dans cette thèse, de nombreuses données réelles sont naturellement aussi des objets tensoriels. Par conséquent, il existe un large éventail d'applications traitant ces objets tensoriels du monde réel. Au-delà de l'analyse des signaux biométrique, de nombreuses tâches de vision par ordinateur et de reconnaissance de formes traitent également avec des objets tensoriels. Ces tâches incluent la reconnaissance d'objets 3D [180], l'analyse d'images médicales [181], la récupération basée sur le contenu [182], l'analyse espace-temps de séquences vidéo pour la reconnaissance des gestes [183] la reconnaissance d'activités [184], l'interaction homme-machine [185] et la reconnaissance faciale hyper-spectrale [186].

En outre, de nombreuses données de transmission sont fréquemment organisées en tenseurs du troisième ordre [187] [188, 189]. Les données de la surveillance des capteurs environnementaux sont souvent organisées en trois modes : heure, lieu et type [187]. Les données de l'analyse des réseaux sociaux sont généralement organisées en trois modes : temps, auteur et mots-clés [187]. Tandis que les données de l'exploration de graphes Web sont généralement organisées en trois modes : source, destination et texte [188]. Les applications basées sur les objets tenseurs discutées ci-dessus sont de plus en plus populaires avec les progrès de la puissance de calcul. Ainsi, il sera intéressant d'étudier l'application des algorithmes d'apprentissage sous-espaces multilinéaires proposés pour résoudre ces problèmes.

Bibliographie

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, pp. 399-458, 2003.
- [2] W.-L. Chao, "Face recognition," GICE, National Taiwan University, 2007.
- [3] S. Lang, "Algebra (graduate texts in mathematics)," Berlin, itd: Springer Verlag, vol. 211, 2002.
- [4] G. Shakhnarovich and B. Moghaddam, "Face recognition in subspaces," in *Handbook of Face Recognition*, ed: Springer, 2005, pp. 141-168.
- [5] M. H. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 377-391, 2006.
- [6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, pp. 71-86, 1991.
- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Yale University New Haven United States* 1997.
- [8] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 328-340, 2005.
- [9] Y. Pang, L. Zhang, Z. Liu, N. Yu, and H. Li, "Neighborhood preserving projections (NPP): a novel linear dimension reduction method," in *International Conference on Intelligent Computing*, 2005, pp. 117-125.
- [10] X. He, S. Yan, Y. Hu, and H.-J. Zhang, "Learning a locality preserving subspace for visual recognition," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, pp. 385-392.
- [11] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis, advances in neural information processing systems," in *Canada: Vancouver*, 2005.
- [12] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear Discriminant Analysis for Face Recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 212-220, 2007.
- [13] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, 2007.
- [14] D. Xu, S. Yan, L. Zhang, H.-J. Zhang, Z. Liu, and H.-Y. Shum, "Concurrent subspaces analysis," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 203-208.
- [15] G. Hua, P. A. Viola, and S. M. Drucker, "Face recognition using discriminatively trained orthogonal rank one tensor projections," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, pp. 1-8.
- [16] Y. Wang and S. Gong, "Tensor discriminant analysis for view-based object recognition," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, pp. 33-36.
- [17] J. Ye, R. Janardan, and Q. Li, "GPCA: an efficient dimension reduction scheme for image compression and retrieval," presented at the *Proceedings of the tenth ACM*

- SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004.
- [18] J. Ye, "Generalized Low Rank Approximations of Matrices," *Machine Learning*, vol. 61, pp. 167-191, 2005/11/01 2005.
 - [19] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-I.
 - [20] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 40-51, 2007.
 - [21] G. Dai and D.-Y. Yeung, "Tensor embedding methods," in *AAAI*, 2006, pp. 330-335.
 - [22] X.-H. Han, X. Qiao, and Y.-W. Chen, "Multilinear supervised neighborhood embedding with local descriptor tensor for face recognition," *IEICE transactions on information and systems*, vol. 94, pp. 158-161, 2011.
 - [23] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," in *null*, 2003, p. 93.
 - [24] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects," *IEEE Transactions on Neural Networks*, vol. 19, pp. 18-39, 2008.
 - [25] H. Wang and N. Ahuja, "Compact representation of multidimensional data using tensor rank-one decomposition," *vectors*, vol. 1, p. 5, 2004.
 - [26] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society vol. 479*: Springer Science & Business Media, 2006.
 - [27] S. Liu and M. Silverman, "A practical guide to biometric security technology," *IT Professional*, vol. 3, pp. 27-32, 2001.
 - [28] D. D. Zhang, *Biometric solutions: For authentication in an e-world vol. 697*: Springer Science & Business Media, 2012.
 - [29] A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Communications of the ACM*, vol. 43, pp. 90-98, 2000.
 - [30] A. Jain, R. Bolle, and S. Pankanti, *BIOMETRIC: Personal identification in networked society*, Kluwer Academic Publishers, 1999.
 - [31] W. Zhang, "Towards non-conventional face recognition: shadow removal and heterogeneous scenario," *Lyon*, 2017.
 - [32] A. Pentland and T. Choudhury, "Face recognition for smart environments," *Computer*, vol. 33, pp. 50-55, 2000.
 - [33] D. Lieberman, J. Tooby, and L. Cosmides, "The architecture of human kin detection," *Nature*, vol. 445, p. 727, 2007.
 - [34] R. Aliradi, A. Belkhir, A. Ouamane, and A. S. Elmaghraby, "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.
 - [35] D. Xu, S. Yan, L. Zhang, S. Lin, H. Zhang, and T. S. Huang, "Reconstruction and Recognition of Tensor-Based Objects With Concurrent Subspaces Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 36-47, 2008.
 - [36] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 721-732, 1997.
 - [37] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 607-626, 2009.

- [38] D. B. Graham and N. M. Allinson, "Face recognition from unfamiliar views: Subspace methods and pose dependency," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 348-353.
- [39] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," 1994.
- [40] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1063-1074, 2003.
- [41] Z.-M. Wang and J.-H. Tao, "Reconstruction of partially occluded face by fast recursive PCA," in *Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on*, 2007, pp. 304-307.
- [42] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu, "Face recognition under occlusions and variant expressions with partial similarity," *IEEE Transactions on Information Forensics and Security*, vol. 4, pp. 217-230, 2009.
- [43] R. Min, A. Hadid, and J.-L. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011, pp. 442-447.
- [44] R. Beveridge and M. Kirby, "Biometrics and Face Recognition," in *IS&T Colloquium, 2005*, p. 25.
- [45] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics: Springer Science & Business Media*, 2007.
- [46] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2518-2525.
- [47] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1253-1278, 2000.
- [48] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r_1, r_2, \dots, r_m) approximation of higher-order tensors," *SIAM journal on Matrix Analysis and Applications*, vol. 21, pp. 1324-1342, 2000.
- [49] B. W. Bader and T. G. Kolda, "Algorithm 862: MATLAB tensor classes for fast algorithm prototyping," *ACM Transactions on Mathematical Software (TOMS)*, vol. 32, pp. 635-653, 2006.
- [50] W. Greub, "Skew symmetry and symmetry in the tensor algebra," in *Multilinear Algebra*, ed: Springer, 1967, pp. 83-94.
- [51] E. Kofidis and P. A. Regalia, "Tensor approximation and signal processing applications," *Contemporary Mathematics*, vol. 280, pp. 103-134, 2001.
- [52] D. Letexier, "Filtrages tensoriels adaptatifs pour la restauration d'images multidimensionnelles," *Aix-Marseille 3*, 2009.
- [53] M. A. O. Vasilescu, "A Multilinear (Tensor) Algebraic Framework for Computer Graphics, Computer Vision and Machine Learning," Citeseer, 2012.
- [54] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognition*, vol. 44, pp. 1540-1551, 2011.
- [55] R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 149-171, 1997/10/01/ 1997.
- [56] H. Lu, *Multilinear subspace learning for face and gait recognition: Citeseer*, 2008.
- [57] L. Haiping, A. C. Kot, and Y. Q. Shi, "Distance-reciprocal distortion measure for binary document images," *IEEE Signal Processing Letters*, vol. 11, pp. 228-231, 2004.
- [58] H. Wang and N. Ahuja, "A Tensor Approximation Approach to Dimensionality Reduction," *Int. J. Comput. Vision*, vol. 76, pp. 217-229, 2008.

- [59] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Advances in neural information processing systems*, 2006, pp. 499-506.
- [60] D. G. Stork, R. O. Duda, P. E. Hart, and D. Stork, "Pattern classification," A Wiley-Interscience Publication, 2001.
- [61] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing vol. 1*: Prentice hall Upper Saddle River, NJ, 2000.
- [62] E. Acar and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 6-20, 2009.
- [63] W. Yong and G. Shaogang, "Tensor Discriminant Analysis for View-based Object Recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 33-36.
- [64] T. Dacheng, L. Xuelong, W. Xindong, and S. Maybank, "Elapsed Time in Human Gait Recognition: A New Approach," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, 2006, pp. II-II.
- [65] G. Hua, P. A. Viola, and S. M. Drucker, "Face Recognition using Discriminatively Trained Orthogonal Rank One Tensor Projections," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
- [66] H. Lu, K. N. Plataniotis, and A. Venetsanopoulos, *Multilinear subspace learning: dimensionality reduction of multidimensional data*: Chapman and Hall/CRC, 2013.
- [67] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*: Wiley-Interscience, 2000.
- [68] I. T. Jolliffe, "Graphical representation of data using principal components," *Principal component analysis*, pp. 78-110, 2002.
- [69] G. Shakhnarovich and B. Moghaddam, "Face Recognition in Subspaces," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., ed London: Springer London, 2011, pp. 19-49.
- [70] K. Fukunaga, "Introduction to statistical pattern classification," ed: Academic Press USA:, 1990.
- [71] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.
- [72] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, pp. 287-314, 1994/04/01/ 1994.
- [73] H. Xiaofei, Y. Shuicheng, H. Yuxiao, P. Niyogi, and Z. Hong-Jiang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 328-340, 2005.
- [74] B. Thompson, *Canonical correlation analysis: Uses and interpretation*: Sage, 1984.
- [75] Y. Ma, P. Niyogi, G. Sapiro, and R. Vidal, "Dimensionality Reduction via Subspace and Submanifold Learning [From the Guest Editors]," *IEEE Signal Processing Magazine*, vol. 28, pp. 14-126, 2011.
- [76] M. H. C. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 377-391, 2006.
- [77] S. Z. Li, C. Zhao, X. Zhu, and Z. Lei, "3D+ 2D face recognition by fusion at both feature and decision levels," in *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, 2005, pp. 43-53.
- [78] R. Chellappa, A. K. Roy-Chowdhury, and S. K. Zhou, "Recognition of humans and their activities using video," *Synthesis Lectures on Image, Video & Multimedia Processing*, vol. 1, pp. 1-173, 2005.
- [79] W. H. Greub, *Multilinear Algebra*. Greub: Springer-Verlag, 1967.

- [80] D. G. Northcott, *Multilinear algebra*: Cambridge University Press, 1984.
- [81] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283-319, 1970.
- [82] P. M. Kroonenberg, *Three-mode principal component analysis: Theory and applications vol. 2*: DSWO press, 1983.
- [83] P. M. Kroonenberg and J. de Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, pp. 69-97, 1980/03/01 1980.
- [84] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279-311, 1966.
- [85] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, vol. 6, pp. 164-189, 1927.
- [86] R. B. Cattell, "'Parallel proportional profiles" and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267-283, 1944.
- [87] R. B. Cattell, "The three basic factor-analytic research designs—their interrelations and derivatives," *Psychological bulletin*, vol. 49, p. 499, 1952.
- [88] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," 1970.
- [89] P. M. Kroonenberg, *Applied multiway data analysis vol. 702*: John Wiley & Sons, 2008.
- [90] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear Analysis of Image Ensembles: TensorFaces," presented at the Proceedings of the 7th European Conference on Computer Vision-Part I, 2002.
- [91] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," in *Object recognition supported by user interaction for service robots*, 2002, pp. 511-514 vol.2.
- [92] Y. Jian, D. Zhang, A. F. Frangi, and Y. Jing-yu, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 131-137, 2004.
- [93] J. Wang, A. Barreto, L. Wang, Y. Chen, N. Rishe, J. Andrian, and M. Adjouadi, "Multilinear principal component analysis for face recognition with fewer features," *Neurocomputing*, vol. 73, pp. 1550-1555, 2010/06/01/ 2010.
- [94] J. Yang, D. Zhang, X. Yong, and J.-y. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognition*, vol. 38, pp. 1125-1129, 2005/07/01/ 2005.
- [95] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1700-1715, 2007.
- [96] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated Multilinear Discriminant Analysis With Regularization and Aggregation for Tensor Object Recognition," *IEEE Transactions on Neural Networks*, vol. 20, pp. 103-123, 2009.
- [97] A. Ouamane, A. Chouchane, E. Boutellaa, M. Belahcene, S. Bourennane, and A. Hadid, "Efficient Tensor-Based 2D+3D Face Verification," *IEEE Transactions on Information Forensics and Security*, vol. 12, pp. 2751-2762, 2017.
- [98] D. Tao, Y. Guo, Y. Li, and X. Gao, "Tensor Rank Preserving Discriminant Analysis for Facial Recognition," *IEEE Transactions on Image Processing*, vol. 27, pp. 325-334, 2018.
- [99] R. Vidal, Y. Ma, and S. Sastry, "Generalized Principal Component Analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1945-1959, 2005.

- [100] Z. Wangmeng, D. Zhang, and W. Kuanquan, "Bidirectional PCA with assembled matrix distance metric for image recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, pp. 863-872, 2006.
- [101] A. A. Mohammed, Q. M. J. Wu, and M. A. Sid-Ahmed, "Application of bidirectional two-dimensional principal component analysis to curvelet feature based face recognition," in *2009 IEEE International Conference on Systems, Man and Cybernetics*, 2009, pp. 4124-4130.
- [102] H. Wang and N. Ahuja, "A Tensor Approximation Approach to Dimensionality Reduction," *International Journal of Computer Vision*, vol. 76, pp. 217-229, 2008/03/01 2008.
- [103] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear principal component analysis through successive variance maximization," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 616-623.
- [104] G. Dai and D.-Y. Yeung, "Tensor embedding methods," presented at the *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, Boston, Massachusetts, 2006.
- [105] Y. Fu and T. S. Huang, "Image Classification Using Correlation Tensor Analysis," *IEEE Transactions on Image Processing*, vol. 17, pp. 226-234, 2008.
- [106] Z. Dazhao, D. Xiufeng, and C. Limin, "Tensor locality preserving projections for face recognition," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 2347-2350.
- [107] S. Yuan, X. Mao, and L. Chen, "Multilinear spatial discriminant analysis for dimensionality reduction," *IEEE Transactions on Image Processing*, vol. 26, pp. 2669-2681, 2017.
- [108] Y. Shuicheng, X. Dong, Y. Qiang, Z. Lei, T. Xiaoou, and Z. Hong-Jiang, "Discriminant analysis with tensor representation," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 526-532 vol. 1.
- [109] K. Fukunaga, *Introduction to statistical pattern recognition*: Elsevier, 2013.
- [110] T. G. Kolda, "Orthogonal tensor decompositions," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, pp. 243-255, 2001.
- [111] W. Songsong, L. Wei, W. Zhisen, and J. Yang, "Local Discriminative Orthogonal Rank-One Tensor Projection for image feature extraction," in *The First Asian Conference on Pattern Recognition*, 2011, pp. 367-371.
- [112] D. Yuan, Y. Liang, L. Yi, Q. Xu, and O. M. Kvalheim, "Uncorrelated linear discriminant analysis (ULDA): A powerful tool for exploration of metabolomics data," *Chemometrics and Intelligent Laboratory Systems*, vol. 93, pp. 70-79, 2008/08/15/ 2008.
- [113] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, pp. 1771-1782, 2000/11/01/ 2000.
- [114] S. Rana, W. Liu, M. Lazarescu, and S. Venkatesh, "A unified tensor framework for face recognition," *Pattern Recognition*, vol. 42, pp. 2850-2862, 2009.
- [115] Y. Liu and Y. Liu, "Tensor distance based multilinear multidimensional scaling for image and video analysis," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 577-580.
- [116] S. Yahia, Y. B. Salem, and M. N. Abdelkrim, "3D face recognition using local binary pattern and grey level co-occurrence matrix," in *Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 2016 17th International Conference on, 2016, pp. 328-338.

- [117] A. Ouamane, A. Chouchane, E. Boutellaa, M. Belahcene, S. Bourennane, and A. Hadid, "Efficient tensor-based 2d+ 3d face verification," *IEEE Transactions on Information Forensics and Security*, vol. 12, pp. 2751-2762, 2017.
- [118] A. Chouchane, A. Ouamane, E. Boutellaa, M. Belahcene, and S. Bourennane, "3D face verification across pose based on euler rotation and tensors," *Multimedia Tools and Applications*, pp. 1-18, 2017.
- [119] T. Mandal, Q. J. Wu, and Y. Yuan, "Curvelet based face recognition via dimension reduction," *Signal Processing*, vol. 89, pp. 2345-2353, 2009.
- [120] G. Zhong and M. Cheriet, "Large margin low rank tensor analysis," *Neural computation*, vol. 26, pp. 761-780, 2014.
- [121] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91.*, IEEE Computer Society Conference on, 1991, pp. 586-591.
- [122] C. E. Thomaz, G. A. Giraldo, J. F. Da Costa, and D. F. Gillies, "A simple and efficient supervised method for spatially weighted PCA in face image analysis," *Technological Reports, Department of Computing in Imperial College, London*, pp. 1-28, 2010.
- [123] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang, and B. Xu, "Generalized discriminant analysis: A matrix exponential approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, pp. 186-197, 2010.
- [124] D. Muti and S. Bourennane, "Survey on tensor signal algebraic filtering," *Signal Processing*, vol. 87, pp. 237-249, 2007.
- [125] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 212-220, 2007.
- [126] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3D face recognition from depth and intensity Gabor features," *Pattern Recognition*, vol. 42, pp. 1895-1905, 2009.
- [127] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Spatially optimized data-level fusion of texture and shape for face recognition," *IEEE Transactions on Image Processing*, vol. 21, pp. 859-872, 2012.
- [128] M. Li and B. Yuan, "A novel statistical linear discriminant analysis for image matrix: two-dimensional fisherfaces," in *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, 2004, pp. 1419-1422.
- [129] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Transactions on Image Processing*, vol. 9, pp. 132-137, 2000.
- [130] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *International conference on image and signal processing*, 2008, pp. 236-243.
- [131] A. Chouchane, M. Belahcène, A. Ouamane, and S. Bourennane, "Evaluation of histograms local features and dimensionality reduction for 3D face verification," *Journal of Information Processing Systems*, pp. 1-21, 2016.
- [132] A. Ouamane, B. Messaoud, A. Guessoum, A. Hadid, and M. Cheriet, "Multi scale multi descriptor local binary features and exponential discriminant analysis for robust face authentication," in *Image Processing (ICIP), 2014 IEEE International Conference on*, 2014, pp. 313-317.
- [133] A. Ouamane, M. Bengherabi, A. Hadid, and M. Cheriet, "Side-information based exponential discriminant analysis for face verification in the wild," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, 2015, pp. 1-6.
- [134] C. Fiche, "Repousser les limites de l'identification faciale en contexte de vidéo-surveillance," *Grenoble*, 2012.

- [135] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on, 2005*, pp. 947-954.
- [136] C. Xu, Y. Wang, T. Tan, and L. Quan, "3D face recognition based on gh shape variation," in *Advances in Biometric Person Authentication*, ed: Springer, 2004, pp. 233-243.
- [137] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *European Workshop on Biometrics and Identity Management, 2008*, pp. 47-56.
- [138] C. C. Queirolo, L. Silva, O. R. Bellon, and M. P. Segundo, "3D face recognition using simulated annealing and the surface interpenetration measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, p. 206, 2010.
- [139] A. Ouamane, E. Boutellaa, M. Bengherabi, A. Taleb-Ahmed, and A. Hadid, "A novel statistical and multiscale local binary feature for 2D and 3D face verification," *Computers & Electrical Engineering*, vol. 62, pp. 68-80, 2017.
- [140] N. Erdogmus and J.-L. Dugelay, "3D Assisted Face Recognition: Dealing With Expression Variations," *IEEE Trans. Information Forensics and Security*, vol. 9, pp. 826-838, 2014.
- [141] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," *Pattern Recognition*, vol. 47, pp. 509-524, 2014.
- [142] F. R. Al-Osaimi, "A novel multi-purpose matching representation of local 3D surfaces: a rotationally invariant, efficient, and highly discriminative approach with an adjustable sensitivity," *IEEE Transactions on Image Processing*, vol. 25, pp. 658-672, 2016.
- [143] S. Elaiwat, M. Bennamoun, F. Boussaïd, and A. El-Sallam, "A curvelet-based approach for textured 3D face recognition," *Pattern Recognition*, vol. 48, pp. 1235-1246, 2015.
- [144] S. Soltanpour and Q. J. Wu, "Multimodal 2D–3D face recognition using local descriptors: pyramidal shape map and structural context," *IET Biometrics*, vol. 6, pp. 27-35, 2016.
- [145] I. A. Kakadiaris, G. Toderici, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S. K. Shah, and T. Theoharis, "3D-2D face recognition with pose and illumination normalization," *Computer Vision and Image Understanding*, vol. 154, pp. 137-151, 2017.
- [146] Y. Guo, R. Wei, and Y. Liu, "Weighted Gradient Feature Extraction Based on Multiscale Sub-Blocks for 3D Facial Recognition in Bimodal Images," *Information*, vol. 9, p. 48, 2018.
- [147] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3D face recognition," *Neurocomputing*, vol. 129, pp. 445-457, 2014.
- [148] L. Best-Rowden, S. Bisht, J. C. Klontz, and A. K. Jain, "Unconstrained face recognition: Establishing baseline human performance via crowdsourcing," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on, 2014*, pp. 1-8.
- [149] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust face recognition for uncontrolled pose and illumination changes," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 43, pp. 149-163, 2013.
- [150] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments," *Expert Systems with Applications*, vol. 47, pp. 23-34, 2016.

- [151] R. Fang, K. D. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," in Image Processing (ICIP), 2010 17th IEEE International Conference on, 2010, pp. 1577-1580.
- [152] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, pp. 331-345, 2014.
- [153] N. Kohli, M. Vatsa, R. Singh, A. Noore, and A. Majumdar, "Hierarchical representation learning for kinship verification," IEEE Transactions on Image Processing, vol. 26, pp. 289-302, 2017.
- [154] S.-J. Wang, S. Yan, J. Yang, C.-G. Zhou, and X. Fu, "A general exponential framework for dimensionality reduction," IEEE Transactions on Image Processing, vol. 23, pp. 920-930, 2014.
- [155] Z. Lai, Y. Xu, J. Yang, J. Tang, and D. Zhang, "Sparse tensor discriminant analysis," IEEE Transactions on Image Processing, vol. 22, pp. 3904-3915, 2013.
- [156] M. Kan, S. Shan, D. Xu, and X. Chen, "Side-Information based Linear Discriminant Analysis for Face Recognition," in BMVC, 2011, pp. 125.1-125.0.
- [157] S. R. Arashloo, "Multiscale binarised statistical image features for symmetric unconstrained face matching," in Electrical Engineering (ICEE), 2014 22nd Iranian Conference on, 2014, pp. 1377-1382.
- [158] J. Kannala and E. Rahtu, "Bsfif: Binarized statistical image features," in Pattern Recognition (ICPR), 2012 21st International Conference on, 2012, pp. 1363-1366.
- [159] A. Hyvärinen, J. Hurri, and P. O. Hoyer, Natural image statistics: a probabilistic approach to early computational vision: Springer.
- [160] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in Asian conference on computer vision, 2010, pp. 709-720.
- [161] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition, 2008.
- [162] H. Yan, J. Lu, W. Deng, and X. Zhou, "Discriminative multimetric learning for kinship verification," IEEE Transactions on Information Forensics and Security, vol. 9, pp. 1169-1178, 2014.
- [163] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding kin relationships in a photo," IEEE Transactions on Multimedia, vol. 14, pp. 1046-1056, 2012.
- [164] X. Qin, X. Tan, and S. Chen, "Tri-subject kinship verification: Understanding the core of a family," IEEE Transactions on Multimedia, vol. 17, pp. 1855-1867, 2015.
- [165] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8.
- [166] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in Workshop on faces in 'real-life' images: Detection, alignment, and recognition, 2008.
- [167] N. Pinto, J. J. DiCarlo, and D. D. Cox, "How far can you get with a modern face recognition test set using only simple features?," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 2591-2598.
- [168] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher Vector Faces in the Wild," in BMVC, 2013, p. 4.
- [169] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt, "Eigen-pep for video face recognition," in Asian Conference on Computer Vision, 2014, pp. 17-33.
- [170] S. R. Arashloo and J. Kittler, "Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features," IEEE Transactions on Information Forensics and Security, vol. 9, pp. 2100-2109, 2014.

- [171] H. Li and G. Hua, "Hierarchical-pep model for real-world face recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4055-4064.
- [172] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust statistical frontalization of human and animal faces," International journal of computer vision, vol. 122, pp. 270-291, 2017.
- [173] J. Lu, J. Hu, and Y.-P. Tan, "Discriminative deep metric learning for face and kinship verification," IEEE Transactions on Image Processing, vol. 26, pp. 4269-4282, 2017.
- [174] H. Yan, J. Lu, and X. Zhou, "Prototype-based discriminative feature learning for kinship verification," IEEE transactions on cybernetics, vol. 45, pp. 2535-2545, 2015.
- [175] Y.-G. Zhao, Z. Song, F. Zheng, and L. Shao, "Learning a Multiple Kernel Similarity Metric for kinship verification," Information Sciences, vol. 430, pp. 247-260, 2018.
- [176] M. Shao, S. Xia, and Y. Fu, "Genealogical face recognition based on ub kinface database," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on, 2011, pp. 60-65.
- [177] N. Kohli, R. Singh, and M. Vatsa, "Self-similarity representation of weber faces for kinship classification," in Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on, 2012, pp. 245-250.
- [178] P. Alirezazadeh, A. Fathi, and F. Abdali-Mohammadi, "Effect of Purposeful Feature Extraction in High-dimensional Kinship Verification Problem," Journal of Computing and Security, vol. 3, pp. 183-191, 2016.
- [179] X. Wu, E. Boutellaa, M. B. López, X. Feng, and A. Hadid, "On the usefulness of color for kinship verification from face images," in Information Forensics and Security (WIFS), 2016 IEEE International Workshop on, 2016, pp. 1-6.
- [180] H. S. Sahambi and K. Khorasani, "A neural-network appearance-based 3-D object recognition using independent component analysis," IEEE Transactions on Neural Networks, vol. 14, pp. 138-149, 2003.
- [181] J. Ruiz-Alzola, C.-F. Westin, S. K. Warfield, C. Alberola, S. Maier, and R. Kikinis, "Nonrigid registration of 3D tensor medical data," Medical image analysis, vol. 6, pp. 143-161, 2002.
- [182] X. He, "Incremental semi-supervised subspace learning for image retrieval," in Proceedings of the 12th annual ACM international conference on Multimedia, 2004, pp. 2-8.
- [183] C. Nolker and H. Ritter, "Visual recognition of continuous hand postures," IEEE Transactions on Neural Networks, vol. 13, pp. 983-994, 2002.
- [184] R. D. Green and L. Guan, "Quantifying and recognizing human movement patterns from monocular video images-part i: a new framework for modeling human motion," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, pp. 179-190, 2004.
- [185] E. Shechtman, Y. Caspi, and M. Irani, "Space-time super-resolution," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 531-545, 2005.
- [186] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatio-spectral information fusion and PLS regression," IEEE Transactions on Image Processing, vol. 24, pp. 1127-1137, 2015.
- [187] C. Faloutsos, T. G. Kolda, and J. Sun, "Mining large time-evolving data using matrix and tensor tools," in ICDM Conference, 2007.
- [188] J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: dynamic tensor analysis," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 374-383.

- [189] J. Sun, Y. Xie, H. Zhang, and C. Faloutsos, "Less is more: Sparse graph mining with compact matrix decomposition," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 1, pp. 6-22, 2008.