



**REPUBLIQUE ALGERIENNE  
DEMOCRATIQUE ET POPULAIRE**



**Ministère de l'Enseignement  
Supérieur et de la recherche scientifique**

**Université Abbès Laghrou-Khenchela**

**Faculté de sciences et technologie**

**Département de génie industriel**

**Mémoire présenté par :**

Mohamedou Mariem

**Pour l'obtention du diplôme de master**

Filière : Automatique

Spécialité : Automatique et informatique industrielle

*Thème*

**Réduction de dimension par l'analyse discriminante linéaire :**

**Application à l'identification de visages**

Encadré par :

*Mme Aourar Salima*

Devant les membres du jury :

Mr. Menadi Abdelkerim : M.C.B, Université Abbès Laghrou, Khenchela, Président

Mr. Mallem Ali: M.C.B, Université Abbès Laghrou, Khenchela, Examineur

Mme : Aourar Salima, M.A.A, Université Abbès Laghrou, Khenchela, Directrice

Année universitaire : 2020 - 2021

## ***Remerciements***

*Je tiens à remercier en premier lieu, les membres du jury d'avoir accepté l'évaluation de ce travail.*

*Mes sincères remerciements vont à l'égard du chef de département de génie industriel à l'université Abbès Laghrour, Mr.Lanani Abderahim ainsi que le chef de la spécialité automatique, Mr.Alouani Fouad, pour leurs travaux permanents, leur disponibilité et leurs conseils qui m'ont guidé durant tout mon cursus universitaire.*

*Ce travail n'aurait pas pu voir le jour sans **Mme. Aourar Salima**, Professeur à l'université Abbès Laghrour, qui m'a accordé l'honneur de m'encadrer durant ce parcours.*

*Merci à mes parents, mes frères et sœurs et mon oncle ... . Je suis bien consciente de la chance que cela représente d'avoir une famille à mes côtés qui me donne sans compter (nuit et jour) et je sais que cela n'a pas de prix.*

## *Dédicaces*

*À ma mère, Khadijatou, mon père, Abdallahi*

*Merci pour, l'amour,*

*le soutien, les sacrifices, la compréhension et les encouragements.*

*Avec toute mon affection et gratitude,*

***Mohamedou Abdallahi Mariem***

## Résumé/Abstract

\*\* Ce présent travail met en évidence la réduction de dimension de données et l'intérêt qu'elle accorde au niveau du traitement d'images. La reconnaissance faciale, étant qu'elle représente un sous champ du traitement d'images, la réduction de dimensionnalité est appliquée dessus. Cependant la technique de réduction de dimension employée pour réaliser ce présent travail est l'analyse discriminante linéaire, celle-ci a prouvé son aptitude à résoudre le problème de grandes dimensions avec une capacité de discrimination élevée qui facilite la classification et l'identification faciale.

Mots-clés : Analyse Discriminante linéaire, Réduction de dimension, ADL, Fisherfaces, Identification de visage, Reconnaissance Faciale

\*\* This present work highlights the dimensionality reduction and the interest it gives to the field of image processing. Facial recognition, being that it represents a subfield of image processing, dimensionality reduction is applied to it. However, the dimensionality reduction technique employed to carry out this present work is linear discriminant analysis, which has proven its ability to solve the large dimension problem with a high discriminating capacity which facilitates classification and facial identification.

- Keywords: Linear discriminant analysis, dimensionality reduction, Fisherfaces, facial recognition, face identification

\*\* يبسط هذا العمل الحالي الضوء على تقليل أبعاد البيانات والاهتمام الذي يعطيه لمستوى معالجة الصور. التعرف على الوجه، نظرًا لأنه يمثل حقلًا فرعيًا من معالجة الصور، يتم تطبيق تقليل الأبعاد عليه. تقنية تقليل الأبعاد المستخدمة لتنفيذ هذا العمل هي التحليل التمييزي الخطي ، والذي أثبت قدرته على حل مشكلة الأبعاد الكبيرة بقدرة تمييزية عالية تسهل التصنيف وتحديد الوجه.

- الكلمات المفتاحية: التحليل الخطي التمييزي ، تقليل الأبعاد ، التعرف على الوجه ، تحديد الوجه

## Liste de notations

- $X$  : Matrice de données
- $N$  : Nombre total d'échantillons dans  $X$
- $W$  : Matrice de transformation
- $n_i$  : Nombre d'échantillons dans  $C_i$
- $\mu_i$  : La moyenne de l' $i$ -nième classe
- $\mu$  : Moyenne totale de tous les échantillons
- $S_{wi}$  : Variance intra-classe ou matrice de dispersion de l' $i$ -nième classe  $C_i$
- $S_{bi}$  : Variance interclasse de l' $i$ -nième classe  $C_i$
- $V$  : Vecteurs propres de  $W$
- $v_i$  :  $i$ -nième vecteur propre
- $x_{ij}$  : L' $i$ -nième échantillon dans la  $j$ -nième classe
- $k$  : La dimension de l'espace dimensionnel inférieur  $V_k$
- $x_i$  :  $i$ -nième échantillon
- $m_i$  : La moyenne de la  $i$ -nième classe après projection
- $m$  : La moyenne totale de toutes les classes après projection
- $S_W$  : Variance intra-classe

$S_B$  : Variance interclasse

$\lambda$  : Matrice de valeurs propres

$M$  : Dimension de  $X$  ou nombre de caractéristiques de  $X$

$V_k$  : L'espace dimensionnel inférieur

$C$  : Nombre total de classe

$\lambda_i$  :  $i$ -ième valeur propre

$C_i$  :  $i$ -ième classe

$Y$  : Projection des données originales

## ***Liste des abréviations***

*ADL: Analyse Discriminante Linéaire*

*FDA: Fisher Discriminant Analysis (Analyse discriminante de Fisher)*

*ACP: Analyse en Composantes Principales*

*ACI: Analyse en Composantes Indépendantes*

*SVM: Support Vector Machine (Machine à vecteur de Support)*

*SSS: Small Sample Size problem (Problème de l'échantillon de petite taille)*

*RF: Reconnaissance Faciale*

*RN: Réseaux de Neurones*

*MDS: Multi Dimensional Scaling*

*LFDA: Local Fisher Discriminant Analysis (Analyse Discriminante locale de Fisher)*

*D-LDA: Direct – Linear Discriminant Analysis (Analyse Discriminante linéaire Directe)*

*LPP: Locality-preserving projection*

# Liste de figures

Figure 1.1 : Nomenclature typique des techniques de réduction de dimensionnalité ...	6
Figure 1.2 : Une taxonomie des techniques de réduction de dimensionnalité	7
Figure 1.3 : ADL, illustration pour le cas bidimensionnel.	10
Figure 1.4 : Projection de la matrice de données sur l'espace dimensionnel inférieur d'ADL ( $V_k$ ).	13
Figure 1.5 : Fonction de densité de probabilité des données projetées	16
Figure 1.6 : Exemples de deux classes non linéairement séparables	18
Figure 2.1 : Processus de reconnaissance faciale	22
Figure 2.2 : différentes méthodes déployées dans un système de reconnaissance faciale	23
Figure 2.3 : Exemple de la courbe CMC	24
Figure 2.4 : Étapes du processus d'identification faciale par ADL	26
Figure 2.5 : différence entre le classement de données par LDA, et PCA	26
Figure 3.1 : Total des images contenues dans la base de données ORL	30
Figure 3.2 : Images des dix premières personnes de la base ORL	31
Figure 3.3: Exemple de changements d'éclairage	31
Figure 3.4 : Exemple de changements des expressions faciales	31
Figure 3.5 : Exemple de changements d'orientations du visage	31
Figure 3.6 : Exemple de changements d'échelle	32
Figure.3.7 : Aperçus sur la fonction <code>perform_lda_Ph</code> utilisée pour la création du model ADL	33
Figure 3.8: Exemple de 3 images redimensionnées.	34
Figure 3.9: Courbe CMC pour l'ensemble d'apprentissage 1.	35
Figure 3.10: Courbe CMC du test réciproque	35
Figure 3.11: Courbe CMC pour l'ensemble d'apprentissage 3	36
Figure 3.12 : Courbe CMC pour l'ensemble d'apprentissage 4	36

# Tables de matières

Remerciement	i
Dédicaces	ii
Résumé/Abstract	iii
ملخص	iii
Abreviations	iv
Notations	v
Liste des Figures	vii
<b>INTRODUCTION GENERALE</b>	<b>1</b>

## Chapitre 1

### Réduction de dimension Par

### Analyse discriminante linéaire LDA

1.1. Introduction	4
1.2. Réduction de dimension	4
1.3. Méthodes de réduction de dimension	5
1.3.1. Méthodes de réduction linéaires	7
1.3.1.1. Analyse en Composantes Principales (ACP)	7
1.3.1.2. Analyse en Composantes Indépendantes (ACI)	8
1.3.1.3. Analyse discriminante linéaire (ADL)	8
1.3.2. Méthodes de réduction non linéaires	8
1.3.2.1. Support Vector Machine (SVM)	8
1.3.2.2. Réseaux de neurones (RN)	9
1.3.2.3. Multi Dimensional Scaling (MDS)	9
1.4. Analyse discriminante linéaire (ADL)	10
1.4.1. Principe de l'analyse discriminante linéaire	10
1.4.2. Algorithme ADL	13
1.4.3. Problèmes principaux d'ADL	17

1.4.3.1. Small Sample Size problem (SSS)	17
1.4.3.2. Problème de linéarité	
1.4.4. Solutions aux problèmes principaux d'ADL	17
1.4.4.1. Solution au Small Sample Size problem (SSS)	18
1.4.4.2. Solution au problème de linéarité	18
1.4.5. Autres variantes d'ADL	19
1.4.5.1. Analyse discriminante de Fisher locale (LFLD)	19
1.4.5.2. Sparse LDA	19
1.5. Conclusion	20
	20

## **Chapitre 2**

### **Identification de visage**

2.1. Introduction	
2.2. Biométrie et reconnaissance de visage	22
2.3. Reconnaissance de visages	22
2.3.1. Différentes phases d'un système de reconnaissance de visages	23
2.3.2. Différentes méthodes de reconnaissance de visages	24
2.3.3. Identification de visage	25
2.3.4. Courbe CMC	26
2.3.5. Algorithme d'identification par ADL	26
2.4. Etat de l'art de l'identification de visages par ADL	27
2.5. Conclusion	28
	29

## **Chapitre 3**

### **Simulations et discussions**

3.1. Introduction	
3.2. Base de données ORL	31
3.3. Algorithme et programmation	31
3.4. Simulations et discussions	34
3.4.1. Aspect matériel et logiciel	35
3.4.2. Simulations	35
3.4.2.1. Ensemble d'apprentissage 1 (200 : 200)	36
3.4.2.2. Ensemble d'apprentissage 2 (200 : 200)	37

3.4.2.3. Ensemble d'apprentissage 3 (280 : 120)	37
3.4.2.4. Ensemble d'apprentissage 4 (360 : 40)	38
3.4.3. Discussions des résultats	38
3.5. Conclusion	38
<b>CONCLUSION GENERALE</b>	39
Bibliographie	42

*Introduction*

*Générale*

## *Introduction générale*

Dans de nombreuses applications d'apprentissage automatique et d'exploration de données, on est souvent confronté à des données de très grande dimension. Une dimension élevée augmente les besoins en temps et en espace pour le traitement des données. De plus, en présence de nombreuses caractéristiques non pertinentes et / ou redondantes, les méthodes d'apprentissage ont tendance à se sur-adapter et à devenir moins interprétables. Un moyen courant de résoudre ce problème est la réduction de la dimensionnalité, qui a attiré beaucoup d'attention dans la communauté d'apprentissage automatique au cours des dernières décennies [1].

La réduction de dimension est l'une des techniques clés de la reconnaissance de formes, elle rassemble de nombreuses méthodes, qui visent toutes à présenter des données à grande échelle dans un espace de faible dimension, d'une manière qu'il capture fidèlement les éléments structurels requis des données [2]. La réduction de dimensionnalité est utilisée à de nombreuses fins, elle est présente comme un outil de visualisation pour présenter des données multi variées sous forme accessible, des méthodes d'extraction de caractéristiques en reconnaissance vocale et des transformations préliminaires appliquées aux données avant d'utiliser l'outil de classification. Cette méthode présente plusieurs avantages évidents comme, la facilitation de la visualisation et de la compréhension de données, la réduction de l'espace de stockage nécessaire ainsi que du temps d'apprentissage et d'utilisation, et l'identification des facteurs pertinents [1].

L'objectif principal de l'algorithme de réduction est d'obtenir une représentation compacte et précise des données pour réduire ou éliminer les composants statistiquement redondants. L'une des méthodes statistiques les plus utilisées pour effectuer la réduction de dimension est l'analyse discriminante linéaire.

L'analyse discriminante linéaire (ADL) ou également, l'analyse discriminante linéaire de Fisher est une méthode supervisée de classification de données et d'extraction de caractéristiques, qui utilise les informations d'étiquettes de différentes classes. Cependant, comme son nom l'indique, l'analyse discriminante de Fisher effectue une discrimination totale entre les différentes classes en maximisant la distance ou la variance interclasse, tout en minimisant la variance intra-calasse qui aboutit à rapprocher les échantillons d'une même classe les uns des autres. Elle a connu un grand succès dans le domaine de la reconnaissance de formes, plus précisément dans la reconnaissance de visage (FR pour Face Recognition), dû à sa capacité de regrouper les éléments appartenant à une même classe tout en éloignant les différentes classes afin de générer une nouvelle classification de données et un taux de reconnaissance élevé.

La reconnaissance de visage est l'un des sous-champs du domaine du traitement d'images. De nos jours, ce dernier devient de plus en plus populaire à des fins de sécurité. De nombreuses techniques ont été développées pour la reconnaissance faciale, dans notre travail, nous employons la technique ADL (Analyse discriminante linéaire) pour la réduction de dimension, qui sera appliquée à la reconnaissance faciale notamment, l'identification de visage.

Ce présent travail est partitionné en trois grands chapitres. Le premier chapitre sera consacré à la réduction de dimension et ses différentes techniques, notamment la méthode ADL qui sera présentée en détails avec quelques algorithmes et exemples numériques. Par la suite, le deuxième chapitre abordera l'identification de visage ainsi que ses méthodes et processus, spécialement l'identification par la méthode ADL. Enfin, dans le dernier chapitre, l'application de la méthode ADL en identification de visages et les différents tests effectués seront présentés. Ces derniers seront suivis par des discussions adéquates des résultats obtenus. Le présent mémoire s'achèvera par une conclusion générale et quelques perspectives futures.

# *CHAPITRE 1*

# *Réduction de dimension Par Analyse discriminante linéaire ADL*

## **1.1. Introduction**

Dans divers domaines, de la recherche scientifique, le développement technologique et la nécessité de superviser des systèmes de plus en plus complexes, nécessitent l'analyse de grandes bases de données (signaux, images, document, etc.). Cependant, si dans cette accumulation de données, nous sommes sûrs que nous avons des informations complètes et utiles, elles risquent de se noyer. Cela pose les problèmes de la structure des données et de l'extraction des connaissances [3].

La réduction de dimensionnalité est une étape indispensable dans le processus de prétraitement des données (filtrage, nettoyage, élimination des valeurs aberrantes, etc.).

Dans ce chapitre, nous allons donner quelques concepts et définitions de base liées à la réduction de dimensionnalité et les différentes méthodes utilisées dans ce contexte. Nous exposons spécifiquement l'algorithme de l'analyse discriminante linéaire (ADL), avec un exemple numérique afin de montrer l'efficacité de cette méthode pour réduire la taille. Nous soulignerons particulièrement la place de l'analyse discriminante linéaire parmi les autres méthodes de réduction de dimension; but de ce travail.

## **1.2. Réduction de dimension**

Fréquemment, les données traitées sont de grande dimension comme des images, des signaux vocaux. Ce problème de grande dimension est notamment posé en machine Learning : il est connu sous le nom de fléau ou malédiction de la dimensionnalité, et survient lors des analyses statistiques [2].

La réduction de dimensionnalité est une technique utilisée pour réduire la complexité de l'analyse de données de grande dimension [3]. L'augmentation du nombre de dimensions augmente également le nombre de variables à analyser. Ces dernières sont alors dissimilaires, ce qui empêche la possibilité de regroupement [2]. Par ailleurs, nous sommes confrontés à deux problèmes :

1- Si le nombre de dimensions dépasse deux, la visualisation des données deviendra ambiguë.

Au fur et à mesure que la dimensionnalité augmente, les données deviennent de plus en plus dispersées dans l'espace.

Par conséquent, la solution à ce problème est de trouver une représentation plus compacte des données pour faciliter leur affichage. Mathématiquement, cela équivaut à trouver une variable aléatoire de dimension  $D$ ,  $\mathbf{x} = (x_1, \dots, x_D)$  et une représentation de dimension inférieure  $\mathbf{s} = (s_1, \dots, s_d)$  (où  $d < D$ ) qui conserve au maximum les informations des données originales [5]. Idéalement la dimensionnalité de la nouvelle représentation est égale au nombre de paramètres requis pour décrire les données observées [6]. Cependant, elle est cruciale dans de nombreux domaines étant donné qu'elle facilite la classification, la visualisation ou encore la compression de données de dimension colossale. Elle est aussi utilisable pour la reconnaissance vocale, et l'analyse exploratoire des données.

### **1.3. Méthodes de réduction de dimension**

De nombreuses méthodologies et techniques sont utilisées pour trouver un nouvel espace de dimension faible d'un ensemble de données, ces techniques réduisent considérablement les dimensions des données d'entrée d'origine. Les méthodes de réduction de dimensionnalité peuvent être de deux types : les techniques *d'extraction* de caractéristiques et celles de la *sélection* de caractéristiques.

L'extraction de caractéristiques est une forme de réduction de la dimensionnalité pour extraire certaines caractéristiques importantes du jeu de données en entrée [4]. Elle a pour but de représenter les échantillons de l'espace de départ dans un nouvel espace de plus faible dimension formé de la combinaison linéaire ou non linéaire des échantillons initiaux. Toutefois, La sélection de caractéristiques choisit les caractéristiques les plus pertinentes selon un critère donné. Ensuite, les données sont analysées après projection dans un espace de représentation composé des caractéristiques les plus pertinentes. Cependant, l'interprétation des caractéristiques extraites est plus délicate que l'interprétation des celles sélectionnées [3].

Il existe deux approches différentes disponibles pour la réduction de dimensionnalité, sont : l'approche supervisée et l'approche non supervisée.

Les méthodes supervisées consistent à déterminer les règles qui permettent de séparer un ensemble d'individus en classes connues à priori, elles utilisent des informations de classes. Les données d'entrée sont étiquetées, ce qui signifie que les sorties sont connues contrairement aux méthodes non supervisées. Celles-ci imposent des limites à l'analyse des données sous certaines contraintes du modèle, également on ne dispose d'aucune information à priori sur le regroupement des données en classes, ce contexte d'apprentissage est dit non supervisé [7] [4] [3] [2].

En apprentissage supervisé, on parle de classification, pour des données qualitatives et de régression pour des données quantitatives; en apprentissage non supervisé, on emploie le terme de *clustering*. La

classification consiste à identifier la catégorie à laquelle appartient une nouvelle observation, à partir de ses caractéristiques. L'algorithme alors utilisé, est appelé *classificateur* ou *classifieur*, l'objectif est notamment la reconnaissance de forme (pattern recognition) [2]

Certaines méthodes d'approche supervisée sont comme : l'analyse discriminante linéaire ADL, réseaux de neurone NN (Neural Network). Cependant, les principaux types d'apprentissage non supervisés sont l'analyse en clusters qui n'utilise pas d'informations d'étiquette. Et de même pour les méthodes: Manifold Learning, ACP (Analyse en Composantes Principales), ACI (Analyse en Composantes Indépendantes), SVD (Singular Value Decomposition), et KPCA (Kernel Principal Component Analysis) ... etc, [2] [4]. La figure ci-dessus, illustre les différentes techniques de réduction de dimensionnalité.

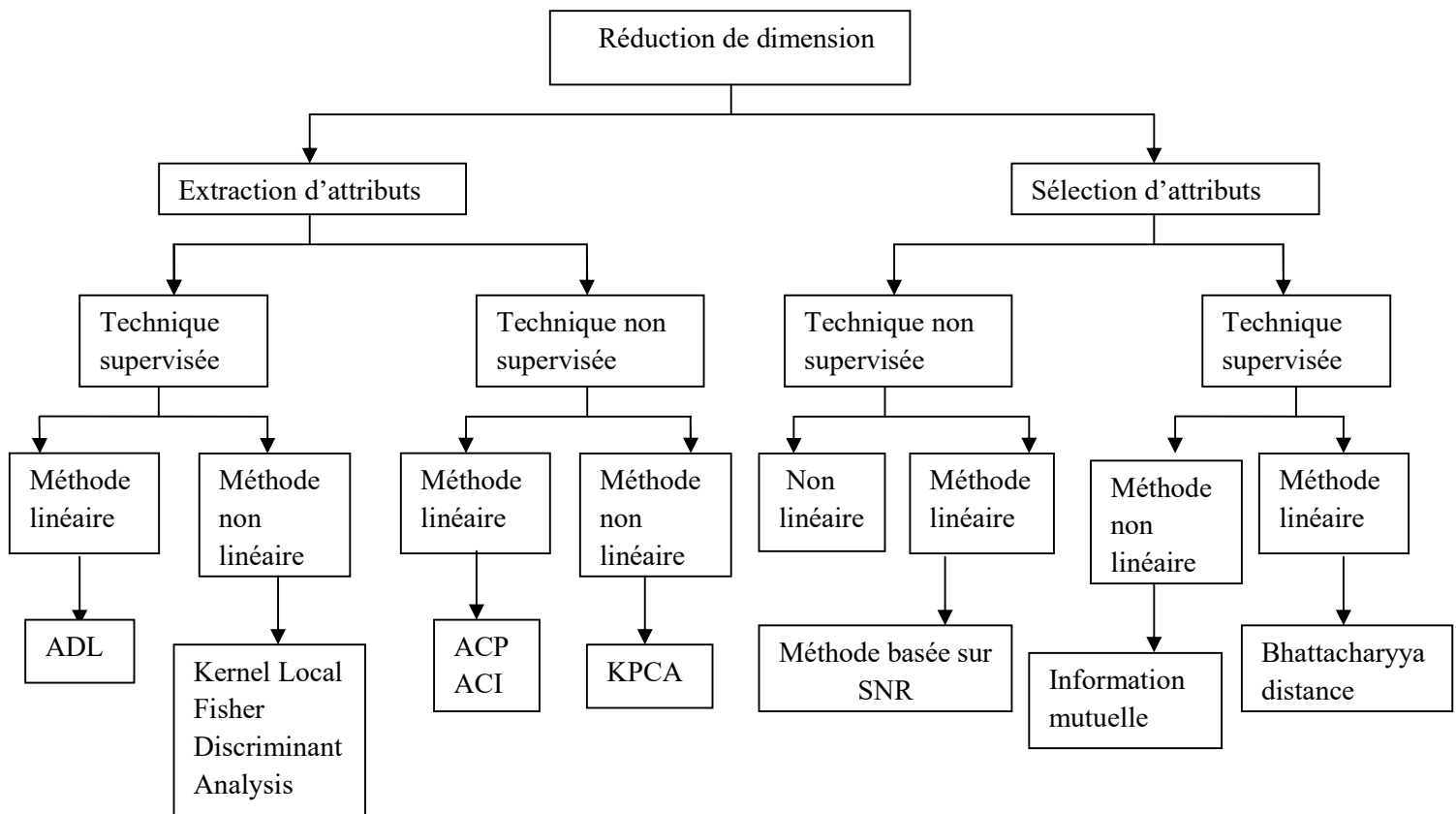
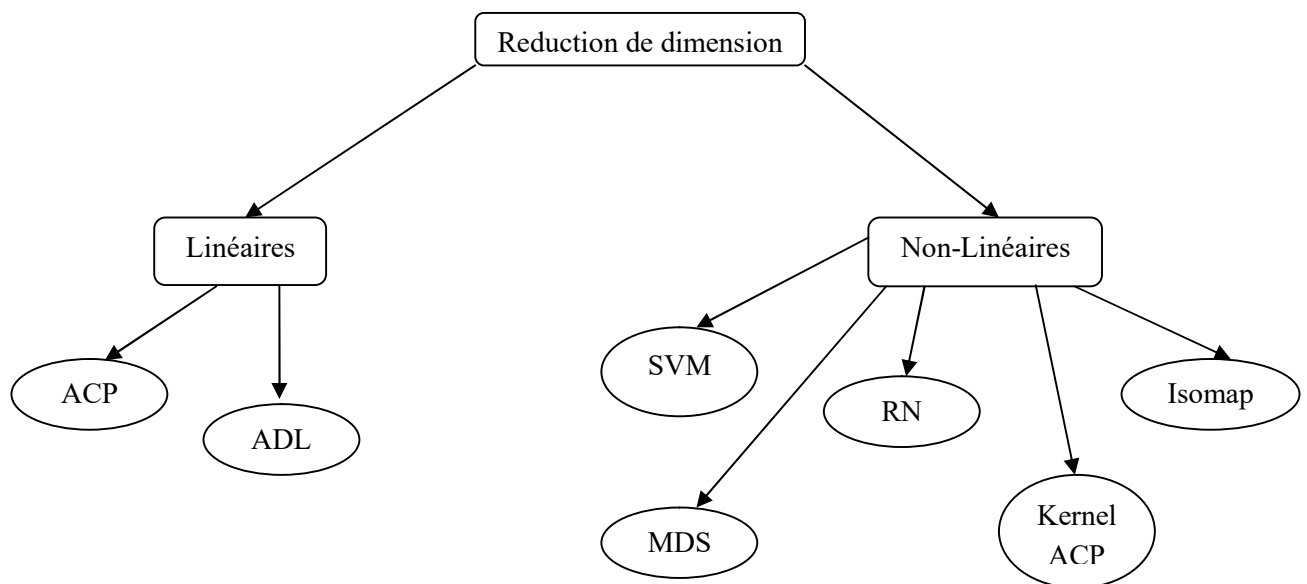


Figure1.1 : Nomenclature typique des techniques de réduction de dimensionnalité.

La réduction de dimensionnalité consiste à résoudre deux types de problèmes différents. L'un est utilisé pour extraire un vecteur de caractéristiques d'un objet original, et l'autre est de réduire la dimensionnalité d'un vecteur de caractéristiques de grande dimension déjà extrait [4].

Un autre type de classification des techniques de réduction de dimensionnalité est celui basé sur le critère linéarité ; la base réduite, est-elle linéaire ou non ? Deux grandes familles sont trouvées dans la

littérature : les méthodes linéaires et les méthodes non linéaires. La figure (1.2) suivante, résume l'ensemble de ces méthodes.



### 1.3.1. Méthodes de réduction linéaires

Dans cette partie, on introduit quelques concepts et définitions des méthodes linéaires les plus fréquemment utilisées: l'Analyse en Composantes Principales (ACP), l'Analyse en Composantes Indépendantes (ICA) et l'Analyse Discriminante Linéaire (LDA ou FLD : *Fisher Linear Discriminant*).

#### 1.3.1.1. Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales (ACP ou PCA pour *Principal Component Analysis*) est une technique de réduction de dimensionnalité largement utilisée pour la reconnaissance et la compression d'images. Elle est également connue sous le nom de transformation de Karhunen-Loeve (KLT) ou projection dans l'espace *propre*. C'est une méthode d'extraction de caractéristiques, non supervisée, elle n'utilise que les mesures, sans tenir compte des étiquettes de classe. Kirby et Sirovich, ont été parmi les premiers à appliquer l'ACP. Cependant, Turk et Pentland ont popularisé l'utilisation de l'ACP pour la reconnaissance faciale dans [8]. Elle a pour objectif d'acquérir une perte minimale d'information, en réduisant la dimension des données d'entrée. Cela s'effectue par la transformation des variables corrélées en nouvelles variables non corrélées également appelées composantes *principales* ou axes *principaux* au sein d'un espace géométrique, ainsi qu'une projection qui maximise la variance tout en minimisant l'erreur quadratique moyenne de la reconstruction [3] [2][8].

### **1.3.1.2. Analyse en Composantes Indépendantes (ACI)**

L'Analyse en Composantes Indépendantes (ou ICA pour *Independent Component Analysis*) a été introduite par Héroult, Jutten et Ans, dans le contexte de la neurophysiologie [3]. C'est une méthode statistique non-supervisée, elle n'utilise donc pas les classes associées aux données. Elle est employée pour transformer un vecteur aléatoire multidimensionnel observé en composantes statistiquement indépendantes les unes des autres, le que possible. L'ICA est un cas particulier de technique de réduction de la redondance, elle devint populaire lors de son utilisation dans le domaine du traitement du signal pour la séparation de sources aveugles. L'ICA d'un vecteur aléatoire consiste à rechercher une transformation linéaire qui minimise la dépendance statistique entre ses composantes, son objectif est de fournir une décomposition et une représentation d'image indépendantes [3] [9].

Comparé aux autres méthodes statistiques, l'ICA fournit une représentation des données plus puissante que l'ACP car cette dernière considère uniquement les moments de second ordre. Tandis que l'ICA prend en compte les statistiques d'ordre supérieur et identifie les composants sources indépendants à partir de leurs mélanges linéaires [9].

### **1.3.1.3. Analyse discriminante linéaire (ADL)**

L'analyse discriminante linéaire s'applique lorsque les classes des individus sont connues et essaie de trouver un sous-espace vectoriel de faible dimension dans lequel les moyennes des attributs des classes sont dispersées par rapport à la covariance interne de chaque classe. Elle maximise la variance inter-classes. Elle projette les données sur ce sous-espace, en évitant aussi des effets collatéraux comme des possibles pertes d'information. Le problème consiste alors à trouver la projection linéaire optimale qui satisfait le critère de séparation (maximisation du rapport des variances inter-classes et intra-classes dans cette projection) [7]. Cette méthode sera détaillée dans la suite du chapitre.

## **1.3.2. Méthodes de réduction non linéaires**

Dans le monde réel, la plupart des données sont sous forme non linéaire. Il est difficile de traiter ces types de données pour une analyse plus approfondie. Il existe de nombreuses méthodes qui permettent de traiter ce type de données non linéaires [8]. Dans cette partie, on évoquera, brièvement, quelques concepts et définitions liées aux méthodes non linéaires de réduction de dimensionnalité, notamment, Support Vector Machine (SVM), Neurones Network (NN) et *Multi Dimensional Scaling (MDS)*.

### **1.3.2.1. Support Vector Machine (SVM)**

Machine à Vecteurs de Support (SVM) a été proposée par V.Vapnik en 1995 [9]. Elle est employée dans divers domaines statistiques (classement, régression, fusion,... etc.) [10]. C'est une technique de classification supervisée qui classe les données en différentes classes et ne considère que les vecteurs de support pour le problème de classification. Chaque ensemble d'enregistrements ou d'instances d'entrée a ses

propres étiquettes de classe. La tâche principale de SVM est de trouver la relation entre l'ensemble de données d'entrée et ses étiquettes de sortie. Généralement, dans le cas non linéaire, les données sont transformées à l'aide de fonctions de transformation non linéaires [11].

Basée sur l'hyper plan, l'idée principale de cette technique consiste à projeter les données de l'espace d'entrée non linéairement séparables, dans un espace de plus faible dimension appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables [12].

### **1.3.2.2. Réseaux de neurones (RN)**

Parmi les techniques non-linéaires de réduction de dimension qui sont couramment utilisées pour la reconnaissance de formes, on trouve celles qui reposent sur un réseau de neurones artificiels (RNA) [10].

Un réseau de neurones (NN) est un système de traitement de l'information qui a été développé comme généralisations des modèles mathématiques inspiré de la physiologie du système nerveux humain. Le (RN) utilise une approche supervisée pour l'analyse. Il est composé d'un grand nombre d'unités de traitement hautement reliées appelées neurones, qui travaillent ensemble pour exécuter une tâche de classification donnée. L'apprentissage des réseaux de neurones est également appelé apprentissage connexionniste en raison de la connexion d'entrée/sortie entre eux. Le réseau de neurones le plus simple contient trois couches : la couche d'entrée, une couche cachée et une couche de sortie [4].

La classification par réseau de neurones contient les étapes suivantes : la première étape est une phase de prétraitement des échantillons d'apprentissage et l'association à chaque échantillon d'apprentissage (entrée du réseau) un vecteur de sortie, puis l'étape d'initialisation (création des couches du réseau), jusqu'à atteindre une certaine erreur minimale (le réseau apprend à bien classifier les échantillons d'apprentissage). On présente ensuite au réseau un nouvel échantillon à identifier (phase de reconnaissance ou de simulation ou d'activation du réseau) qui sera finalement affecté à une classe donnée [13].

### **1.3.2.3. Multi Dimensional Scaling (MDS)**

La méthode *Multi Dimensional Scaling* (MDS) ou encore, de positionnement multidimensionnel (PMD), est la collection de techniques non linéaires pour transformer les données de grande dimension en données de faible dimension. L'erreur entre la distance par paire entre les données de faible dimension et les données de grande dimension est exprimée dans la fonction de contrainte. Les exemples de fonctions de contrainte sont la fonction de contrainte brute et la fonction de coût Sommon [11].

MDS est une transformation de caractéristiques explicite qui trouve directement l'ensemble de données réduit au lieu de la matrice de transformation [14]. La minimisation de la fonction de contrainte réduit l'erreur ce qui conduit à améliorer les performances du système. La méthode du gradient conjugué et la méthode Pseudo-Newton sont quelques-unes des méthodes pour réduire la fonction de contrainte.

## 1.4. Analyse discriminante linéaire (ADL)

Développé à l'origine en 1936 par *R.A. Fisher* (1890-1962) [14], l'analyse discriminante linéaire (ADL) est une méthode d'apprentissage sous-espace supervisé. Elle est basée sur le critère de Fisher, qui joue un rôle important dans la réduction de dimension [16].

ADL a connue un grand succès dans la reconnaissance faciale, elle est connue souvent sous le nom de Fisher Discriminant Analysis (FDA); l'analyse discriminante de Fisher. Elle a pour but de maximiser la distance entre les classes tout en minimisant la variance intra-classes. Cette méthode vise à augmenter la séparabilité entre les différentes classes tout en rapprochant considérablement les échantillons d'une même classe [17].

L'analyse discriminante est une méthode classique de classification qui a résisté à l'épreuve du temps. Elle produit souvent des modèles dont la précision approche des méthodes modernes plus complexes [18]. Nous présentons dans ce qui suit le principe de calcul et les opérations mathématiques pour l'application de cette méthode.

### 1.4.1. Principe de l'analyse discriminante linéaire

L'idée est de trouver une projection des échantillons, sur une droite qui sépare le mieux possible les classes. L'ADL a donc pour objectif de trouver une transformation linéaire  $W \in R^{d \times m}$  qui est choisie comme étant la projection linéaire de l'espace de départ de dimension  $d$  sur un espace de dimension  $m$  plus faible tel que  $m \ll d$ . Dans ce nouvel espace, la dispersion entre les classes est maximisée tandis que la dispersion intra-classe est minimisée [11]. Autrement dit, la distribution de données d'une même classe doit être compacte, ce qui revient à minimiser la dispersion autour de la moyenne, ainsi qu'à la maximiser entre les classes [7]. Ce principe est illustré par la figure ci après.

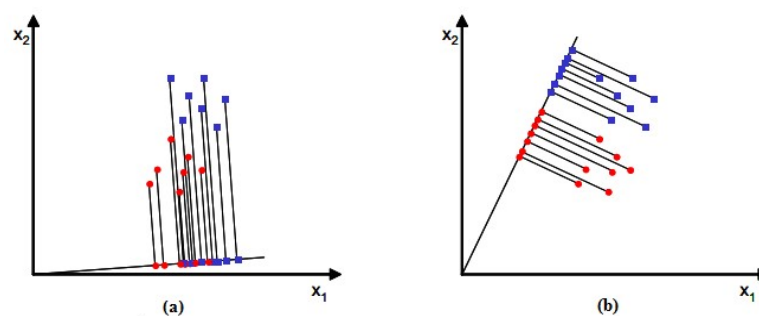


Figure 1.3: ADL, illustration pour le cas bidimensionnel. [19]

a : Projection non ADL    b : Projection ADL

Afin de réaliser cet objectif, trois étapes doivent être suivies : la première étape consiste à calculer la séparabilité entre les différentes classes, qui est appelée la variance interclasses ou encore la matrice de dispersion inter-classes. L'étape suivante est le calcul de la variance intra-classes ou également la matrice de

dispersion intra-classes. La dernière étape consiste à construire l'espace de dimension inférieur, qui maximise la variance inter-classes et minimise la variance intra-classes [20].

**a. Calcul de la matrice inter-classes**

Étant donné la matrice de données d'origine  $X = \{x_1, x_2, \dots, x_N\}$ , partitionnée en ( $c$ ) classes, où  $x_i$  représente le  $i^{\text{ème}}$  échantillon, modèle ou observation,  $n_i$  représente le nombre d'échantillons de la  $i^{\text{ème}}$  classe.  $N$  est le nombre total d'échantillons. Chaque échantillon est représenté par  $M$  caractéristiques ( $x_i \in R^M$ ). En d'autres termes, chaque échantillon est représenté comme un point dans l'espace dimensionnel  $M$  [19]. La solution proposée par Fisher est de maximiser une fonction exprimée par la différence entre les moyennes, normalisée par une mesure de la dispersion intra-classes [20].

La variance inter-classes de la  $i^{\text{ème}}$  classe ( $S_{bi}$ ) représente la distance entre la moyenne de la  $i^{\text{ème}}$  classe ( $\mu_i$ ) et la moyenne totale ( $\mu$ ) [18]. Elle est donnée comme suit :

$$S_{bi} = \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^T \dots\dots\dots (1)$$

La matrice de dispersion entre les classes ( $S_B$ ) mesure le montant de dispersion entre les classes. Elle représente la somme des matrices de covariance ( $S_{bi}$ ) [17]. Elle se calcule comme suit :

$$S_B = \sum_{i=1}^c n_i S_{bi} \dots\dots\dots (2)$$

$\mu_i (1 \times M)$  représente la moyenne de la  $i^{\text{ème}}$  classe et elle est donnée par :

$$\mu_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i \dots\dots\dots (3)$$

Et  $\mu (1 \times M)$  est la moyenne totale de toutes les classes et elle est calculée comme suit :

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i = \sum_{i=1}^c \frac{n_i}{N} \mu_i \dots\dots\dots (4)$$

**b. Calcul de la matrice intra-classes**

La matrice de dispersion intra-classe mesure la quantité éparpillée entre les éléments de la même classe. Pour la  $i^{\text{ème}}$  classe, la matrice de dispersion ( $S_{wi}$ ) est calculée comme la somme des matrices de covariance des échantillons centrées dans cette classe [18] [20]. Pour chaque classe, nous définissons sa dispersion comme :

$$S_{wi} = \sum_{x \in x_i} (x - \mu_i)(x - \mu_i)^T \dots\dots\dots (5)$$

La matrice de dispersion intra-classe ( $S_W$ ) est la somme de toutes les matrices de dispersion :

$$S_W = \sum_{i=1}^c S_{wi} \dots\dots\dots (6)$$

C est le nombre de classes.

**c. Construction de l'espace dimensionnel inferieur**

L'analyse discriminante linéaire de Fisher est définie comme la fonction linéaire,  $y = W^T x$  qui maximise le critère de Fisher nommé  $J(W)$ . Par conséquent, nous rechercherons une projection où les exemples d'une même classe sont projetés très proches les uns des autres et, en même temps, les moyennes projetées sont aussi éloignées que possible [8] [9] [19]. Cependant, la matrice de transformation ( $W$ ) d'ADL est donnée par l'équation:

$$J(W) = \mathbf{arg\ max} \frac{W^T S_B W}{W^T S_w W} \dots\dots\dots (7)$$

Cette formule peut être reformulée comme dans l'équation ci-dessus :

$$S_w W = \lambda S_B W \dots\dots\dots (8)$$

Ces vecteurs propres représentent les directions du nouvel espace et les valeurs propres correspondantes représentent le facteur d'échelle, la longueur ou la grandeur des vecteurs propres [20]. Ainsi, chaque vecteur propre représente un axe de l'espace ADL, et la valeur propre associée représente la robustesse de ce vecteur propre.

Les vecteurs propres doivent être ordonnés par ordre décroissant des valeurs propres correspondantes afin de choisir ceux qui sont à retenir et ceux qui sont à négliger. La matrice de transformation d'ADL sera constituée par les premiers vecteurs propres significatifs [20] [17].

Ainsi, les vecteurs propres avec les  $k$  valeurs propres les plus élevées sont utilisés pour construire un espace dimensionnel inférieur ( $V_k$ ). Tandis que les autres vecteurs propres ( $\{V_{k+1}, V_{k+2}, \dots, V_M\}$ ) sont négligés [3] [20].

La figure ci-dessus illustre l'espace dimensionnel inférieur calculé. La dimension de la matrice de données d'origine ( $X \in R^{N \times M}$ ) est réduite en la projetant sur l'espace dimensionnel inférieur d'ADL ( $V_k \in R^{M \times k}$ ); indiqué par l'équation  $Y = XV_k$ . La dimension des données après projection est  $k$ . Par conséquent, les fonctionnalités  $M - k$  sont ignorées ou supprimées de chaque échantillon. Ainsi, chaque échantillon ( $x_i$ ) vu comme étant un point d'un espace à  $M$  dimensions, est représenté dans un espace à  $k$  dimensions en le projetant sur l'espace dimensionnel inférieur ( $V_k$ ) comme suit,  $y_i = x_i V_k$  [20]

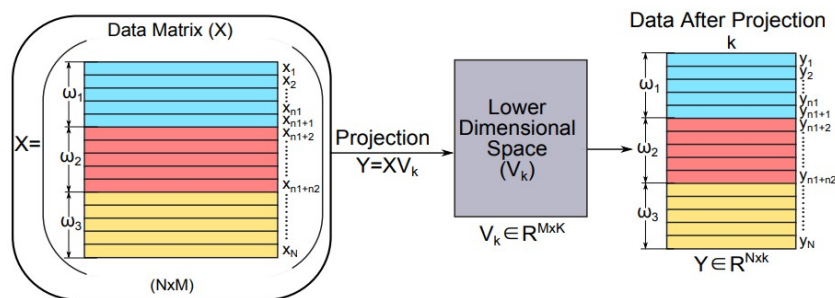


Fig.1.4 : Projection de la matrice de données sur l'espace dimensionnel inférieur d'ADL ( $V_k$ ). [20].

### 1.4.2. Algorithme ADL

Deux versions d'algorithme ADL existent : la méthode indépendante des classes (*Class-Independent method*) et la méthode dépendante des classes (*Class-Dependant method*). Cette dernière calcule l'espace LDA séparément pour chaque classe comme suit :  $W_i = S_{w_i}^{-1} S_B$  où  $W_i$  représente la matrice de transformation pour la  $i^{\text{ème}}$  classe. De même les vecteurs propres et les valeurs propres sont calculés séparément pour chaque matrice de transformation [21] [20] [19]. Par conséquent, les échantillons de chaque classe sont projetés sur leurs vecteurs propres correspondants.

Par contre, dans la méthode de *class independant* ; telle est la méthode standard d'ADL, l'espace dimensionnel inférieur est calculé pour toutes les classes. Ainsi, la matrice de transformation est calculée pour toutes les classes au même temps  $W = S_W^{-1} S_B$  où  $W_i$  et les échantillons de toutes les classes sont projetés sur les vecteurs propres sélectionnés [20] [22]. Les étapes détaillées de l'algorithme de cette méthode sont présentées dans cette partie.

#### a. Algorithme (ADL): méthode Class independant

Étant donné un ensemble de  $N$  échantillons  $[x_i]_{i=1}^N$ , dont chacun est représenté comme une ligne de longueur  $M$  et  $X(N \times M)$  donné par :

$$\begin{bmatrix} x(1,1) & x(1,2) & \cdots & x(1,M) \\ \vdots & \vdots & \vdots & \vdots \\ x(N,1) & x(N,2) & \cdots & x(N,M) \end{bmatrix}$$

- 1 : Calcul de la moyenne de chaque classe  $\mu_i$  ( $1 \times M$ ).
- 2: Calcul de la moyenne totale de toutes les données  $\mu$  ( $1 \times M$ ).
- 3: Calcul de la matrice interclasses  $S_B$  ( $M \times M$ ).
- 4: Calcul de la matrice intra-classe  $S_W$  ( $1 \times M$ ).
- 5: Calcul de la matrice  $W$  qui maximise la formule de Fisher  $W = S_W^{-1} S_B$ .
- 6 : Calcul des valeurs propres ( $\lambda$ ) et les vecteurs propres ( $V$ ) de la matrice  $W$ .
- 7: Tri des vecteurs propres par ordre décroissant en fonction de leurs valeurs propres correspondantes.
- 8: Utilisation des  $k$  premiers vecteurs propres pour construire l'espace dimensionnel inférieur ( $V_K$ ).
- 9 : Projection des échantillons originaux ( $X$ ) sur l'espace dimensionnel inférieur d'ADL.

### b. Exemple numérique

Dans cet exemple, l'algorithme précédant pour discriminer seulement deux classes différentes. Nous montrons ainsi, comment ses différentes étapes sont calculées.

Soit deux classes différentes,  $C_1$  ( $5 \times 2$ ) et  $C_2$  ( $6 \times 2$ ) ayant respectivement ( $n_1 = 5$ ) et ( $n_2 = 6$ ) échantillons chacune. Chaque échantillon dans les deux classes est représenté par deux caractéristiques ( $M = 2$ ) comme suit:

$$C_1 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{bmatrix} \quad \text{Et} \quad C_2 = \begin{bmatrix} 4 & 2 \\ 5 & 0 \\ 5 & 2 \\ 3 & 2 \\ 5 & 3 \\ 6 & 3 \end{bmatrix}$$

Pour calculer l'espace dimensionnel inférieur de LDA, la moyenne de chaque classe  $\mu_j$  est d'abord calculée. La moyenne totale  $\mu$  ( $1 \times 2$ ) est ensuite calculée, qui représente la moyenne de toutes les moyennes de toutes les classes.

- 1: Les valeurs de la moyenne  $\mu_i$  ( $1 \times M$ ) de chaque classe sont obtenues comme suit :

$$\mu_1 = [3.00 \quad 3.60], \quad \mu_2 = [4.67 \quad 2.00]$$

- 2: La moyenne totale est alors indiquée ci-dessous :

$$\mu = \left[ \frac{5}{11} \mu_1 \quad \frac{6}{11} \mu_2 \right] = [3.91 \quad 2.727]$$

**3:** La variance inter-classes de chaque classe ( $S_{bi} (2 \times 2)$ ) et la variance totale interclasses ( $S_B (2 \times 2)$ ) sont calculées. Les valeurs de la variance interclasses de la première classe ( $S_{b1}$ ) sont égales à :

$$S_{b1} = n_1(\mu_1 - \mu)^T(\mu_1 - \mu) = 5[-0.91 \ 0.87]^T[-0.91 \ 0.87] = \begin{bmatrix} 4.13 & -3.97 \\ -3.97 & 3.81 \end{bmatrix}$$

De même,  $S_{b2}$  est égale à:

$$S_{b2} = \begin{bmatrix} 3.44 & -3.31 \\ -3.31 & 3.17 \end{bmatrix}$$

La variance totale interclasse est calculée comme suit :

$$S_B = S_{b1} + S_{b2} = \begin{bmatrix} 4.13 & -3.97 \\ -3.97 & 3.81 \end{bmatrix} + \begin{bmatrix} 3.44 & -3.31 \\ -3.31 & 3.17 \end{bmatrix} = \begin{bmatrix} 7.58 & -7.27 \\ -7.27 & 6.98 \end{bmatrix}$$

**4:** Pour calculer la matrice intra-classes, on doit d'abord soustraire la moyenne de chaque classe de chaque échantillon dans cette classe et cette étape est appelée centrage des données (*mean-centering data*). Elle est calculée par,  $d_i = C_i - \mu_i$  où  $d_i$  représente les données de centrage de la classe  $C_i$ . Les valeurs de  $d_1$  et  $d_2$  sont les suivantes:

$$d_1 = \begin{bmatrix} -2.00 & -1.60 \\ -1.00 & -0.60 \\ 0.00 & -0.60 \\ 1.00 & 1.40 \\ 2.00 & 1.40 \end{bmatrix} \quad \text{et} \quad d_2 = \begin{bmatrix} -0.67 & 0.00 \\ 0.33 & -2.00 \\ 0.33 & 0.00 \\ -1.67 & 0.00 \\ 0.33 & 1.00 \\ 1.33 & 1.00 \end{bmatrix}$$

Après avoir centré les données, la variance intra-classe pour chaque classe ( $S_{wi} (2 \times 2)$ ) est calculée comme suit :

$$S_{wj} = d_j^T * d_j = \sum_{i=1}^{n_j} (x_{ij} - \mu_j)^T (x_{ij} - \mu_j), \text{ où } x_{ij} \text{ représente le } i^{\text{ème}} \text{ échantillon de la } j^{\text{ème}} \text{ classe.}$$

Les valeurs de la matrice intra-classes pour chaque classe et la matrice totale intra-classes ( $S_W (2 \times 2)$ ) sont les suivantes :

$$S_{w1} = \begin{bmatrix} 10.00 & 8.00 \\ 8.00 & 7.20 \end{bmatrix}, \quad S_{w2} = \begin{bmatrix} 5.33 & 1.00 \\ 1.00 & 6.00 \end{bmatrix}, \quad S_W = \begin{bmatrix} 15.33 & 9.00 \\ 9.00 & 13.20 \end{bmatrix}$$

**5:** La matrice de transformation ( $W$ ) est calculée, et les valeurs de ( $S_W^{-1}$ ) et ( $W$ ) sont les suivantes:

$$S_W^{-1} = \begin{bmatrix} 0.11 & -0.07 \\ -0.07 & 0.13 \end{bmatrix} \quad \text{et} \quad W = \begin{bmatrix} 1.36 & -1.31 \\ -1.48 & 1.42 \end{bmatrix}$$

**6:** Alors, les valeurs propres ( $\lambda (2 \times 2)$ ) et les vecteurs propres ( $V (2 \times 2)$ ) de  $W$  sont les suivants:

$$\lambda = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 2.78 \end{bmatrix} \quad \text{et} \quad V = \begin{bmatrix} -0.69 & 0.68 \\ -0.72 & -0.74 \end{bmatrix}$$

6: D'après les résultats ci-dessus, on peut remarquer que, le deuxième vecteur propre ( $V_2$ ) a une valeur propre correspondante plus que le premier ( $V_1$ ), ce qui reflète que, le deuxième vecteur propre est plus robuste que le premier;

7: Par conséquent, le vecteur ( $V_2$ ) est choisi pour construire l'espace dimensionnel inférieur.

8: Les données originales sont projetées sur l'espace dimensionnel inférieur par,  $y_i = C_i V_2$  où  $y_i (n_i \times 1)$  représente les données après projection de la  $i^{ème}$  classe, et ses valeurs seront les suivantes :

$$y_1 = C_1 V_2 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} 0.68 \\ -0.74 \end{bmatrix} = \begin{bmatrix} -0.79 \\ -0.85 \\ -0.18 \\ -0.97 \\ -0.29 \end{bmatrix}$$

De même,  $y_2$  est comme suit :  $y_2 = C_2 V_2 = \begin{bmatrix} 1.24 \\ 3.39 \\ 1.92 \\ 0.56 \\ 1.18 \\ 1.86 \end{bmatrix}$

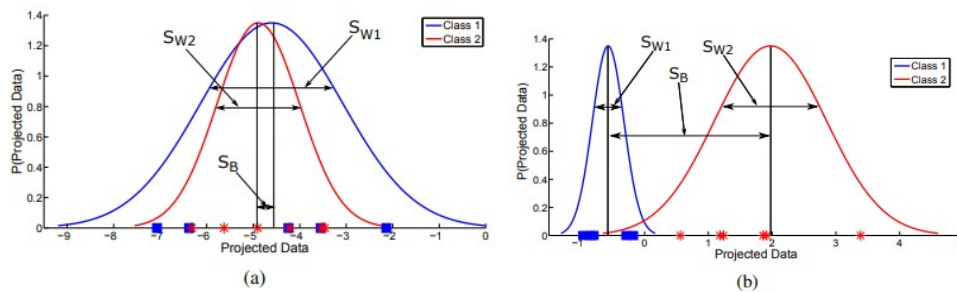


Figure 1.5: Fonction de densité de probabilité des données projetées de l'exemple précédent [20], (a) les données projetées sur  $V_1$ , (b) les données projetées sur  $V_2$ .

La figure 1.5 ci-dessus, illustre de *fonction de densité de probabilité* des données projetées ( $y_i$ ) sur les deux vecteurs propres ( $V_1$  et  $V_2$ ). Une comparaison des deux vecteurs propres révèle ce qui suit:

- \* Le deuxième vecteur propre maximise la variance entre les classes plus que le premier, on remarque que les données de chaque classe sont complètement discriminées lorsqu'elles sont projetées sur le deuxième vecteur propre ; Figure 1.5.b que le premier ; Figure 1.5.a.

- \* La variance intra-classes des deux classes est minimisée lorsque les données sont projetées sur le deuxième vecteur propre ; Figure 1.5.b. La variance intra-classes de la première classe est faible par rapport à la Figure 1.5.a [20].

### 1.4.3. Problèmes principaux d'ADL

Bien que LDA soit l'une des techniques de réduction de données les plus courantes, elle souffre de deux problèmes principaux : les problèmes de petite taille d'échantillon (SSS), et de linéarité [20].

#### 1.4.3.1. Small Sample Size problem (SSS)

C'est le problème de petites tailles des échantillons. En effet, le nombre d'échantillons d'apprentissage disponibles est généralement beaucoup plus petit que la dimensionnalité de l'espace de l'échantillon.

La petite taille d'échantillon (SSS), la singularité ou le problème de sous-échantillonné est l'un des gros problèmes de la technique LDA. Ce problème est rencontré lors de tâches de classification de motifs de grande dimension ou d'un faible nombre d'échantillons d'apprentissage disponibles pour chaque classe par rapport à la dimensionnalité de l'espace d'échantillonnage [20]. En effet La théorie de l'apprentissage statistique affirme que la difficulté d'un problème d'estimation (par exemple pour  $S_B$  et  $S_W$ ) augmente considérablement avec la dimensionnalité  $J$  de l'espace échantillon.

Le problème SSS survient lorsque le  $S_W$  est singulier. La borne supérieure du rang de  $S_W$  est  $N - c$  tandis que la dimension de  $S_W$  est  $M \times M$ . Ainsi, dans la plupart des cas  $M \gg N - c$  ce qui conduit au problème SSS. Par exemple, dans les applications de reconnaissance faciale, la taille de l'image du visage peut atteindre  $100 \times 100 = 10\,000$  pixels ce qui représente des caractéristiques de grande dimension. De plus, le nombre ( $C_i$ ) d'exemples par classe disponibles pour l'apprentissage ne dépasse pas dix dans la plupart des cas, ce qui conduit à un problème de singularité, ayant une influence significative sur la conception et les performances de ce système de reconnaissance [21] [23].

#### 1.4.3.2. Problème de linéarité

L'analyse discriminante linéaire est également confrontée à un autre problème dit de linéarité. La technique LDA est utilisée pour trouver une transformation linéaire qui discrimine les différentes classes. Cependant, si les classes sont séparables de manière non linéaire, elle devient incapable de discriminer entre les classes. Par conséquent, Elle ne parvient pas à trouver l'espace LDA. Cette limitation est rencontrée lorsque les informations discriminantes ne sont pas dans les moyennes des classes.

La Figure (1.6) montre comment l'information discriminante n'existe pas dans la moyenne, mais dans la variance des données. C'est parce que les moyennes des deux classes sont égales. L'interprétation mathématique révèle que  $S_B$  et  $W$  seront nuls si les moyennes des classes sont approximativement égales. Par conséquent, l'espace LDA ne peut pas être calculé. Néanmoins, il existe une solution à ce problème, basée sur le concept de transformation, connu sous le nom de méthodes ou fonctions du noyau (*Kernel methods*) [20].

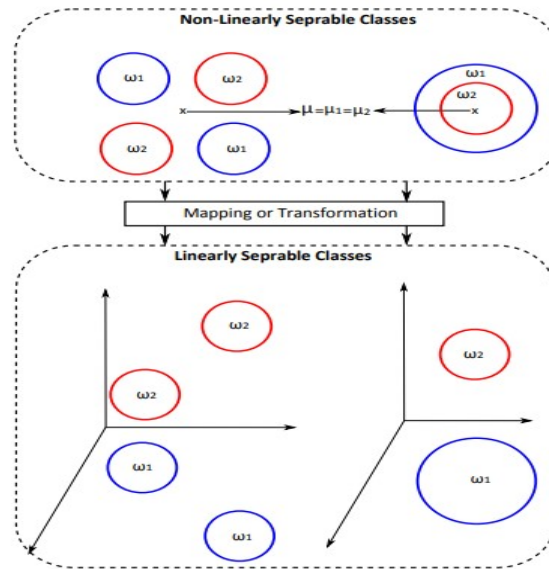


Figure. 1.6 : Exemples de deux classes non linéairement séparables.  
En haut : les deux classes sont non séparables. En bas transformation résout ce problème et les deux classes sont linéairement séparables [20].

#### 1.4.4. Solutions aux problèmes principaux d'ADL

##### 1.4.4.1. Solution au Small Sample Size problem (SSS)

Au cours des trois dernières décennies, des efforts de recherche considérables ont été déployés pour traiter le problème SSS lié à divers problèmes tels l'extraction de caractéristiques, la sélection de caractéristiques et la conception de classificateurs [23]. De nombreuses solutions sont proposées dans la littérature ; chacun a ses avantages et ses inconvénients [20]. Brièvement, on en cite :

- Une solution basée sur l'approche de sous-espace, telle que celle suivie dans le développement de la méthode *Fisherfaces*, où PCA est d'abord utilisé comme étape de prétraitement pour supprimer l'espace nul de  $S_W$ , puis LDA est effectué dans le sous-espace PCA de dimension inférieure [23]. Dans cette méthode, un espace intermédiaire non singulier est obtenu pour réduire la dimension des données d'origine pour qu'elle soit égale au rang de  $S_W$  ; par conséquent,  $S_W$  devient de rang complet, puis  $S_W$  peut être inversé et le problème SSS est résolu. Par exemple, Belhumeur et al. dans [17] ont utilisé l'ACP, pour réduire les dimensions de l'espace d'origine à  $N - c$ . Cependant, la perte de certaines informations discriminantes est un inconvénient courant, associé à l'utilisation de cette méthode [20].
- Une solution, appelée méthode LDA directe (DLDA): La LDA directe est l'une des techniques célèbres utilisées pour résoudre le problème SSS. Elle comporte deux étapes principales [20]. La première étape, consiste à calculer la matrice de transformation  $W$ , pour transformer les données d'apprentissage dans l'espace de plage de  $S_B$ . Dans la deuxième étape, la dimensionnalité des données transformées est encore transformée en utilisant des matrices de régulation. La prémisse de

base derrière l'approche DLDA est que l'espace nul de  $S_W$  contient des informations discriminantes significatives si la projection de  $S_B$  n'est pas nulle dans cette direction [23]. L'avantage de cette méthode est qu'aucune caractéristique discriminante n'est négligée comme dans la technique PCA+LDA [20].

- L'algorithme de DFLDA (DIRECT FRACTIONAL LDA) : une technique qui traite également le problème SSS d'une façon efficace et rentable, tout en offrant des performances supérieures à celles des techniques de reconnaissance de visages classiques [24].

#### **1.4.4.2. Solution au problème de linéarité**

Plusieurs études basées sur les méthodes du noyau (kernel methods), ont montré l'aptitude de ces techniques à résoudre le problème de linéarité souvent rencontré en ADL.

Par exemple, l'analyse discriminante linéaire de Kernel Fisher (KFLDA) a été proposée pour la classification binaire non linéaire, elle s'agit d'une méthode hybride de l'analyse discriminante linéaire classique de Fisher et d'une machine à noyau. Les résultats expérimentaux ont montré que la KFLDA fonctionne mieux en termes d'erreur de prédiction que les machines à vecteurs de support populaires et est un concurrent sérieux de ces dernières. Cependant, la justification statistique de cette méthode est très limitée [25].

#### **1.4.5. Autres variantes d'ADL**

Dans cette partie, on cite quelques variantes d'ADL, notamment, ADL locale de Fisher (*LFDA*) et *sparse* ADL (*S-LDA*).

##### **1.4.5.1. Analyse discriminante de Fisher locale (LFLD)**

L'analyse discriminante locale de Fisher (*LFDA*), est une variante localisée de l'analyse discriminante de Fisher. Elle prend en compte la structure locale des données afin que les données multimodales puissent être intégrées de manière appropriée [26].

Dans ce qui suit, les explications sont entre autres, inspirées de l'article de : Masashi Sugiyama. [26]

Local Fisher Discriminant Analysis combine les idées de la FDA et de la LPP (*Locality-preserving projection*). Celle-ci est une méthode de réduction de dimension non supervisée, qui ne prend pas les informations d'étiquette en compte. Cependant, la LPP conserve les paires de données à proximité dans l'espace original, proche de l'espace d'intégration, par lequel les données multimodales peuvent être intégrées sans perdre leur structure: La combinaison des deux méthodes précédentes qu'exerce la méthode LFDA conduit à :

- \* la séparabilité entre les classes est maximisée tandis que la structure locale au sein de la classe est préservée.

\* En raison de la propriété de préservation de la structure locale, les données étiquetées multimodales peuvent être efficacement intégrées par LFDA.

### **1.4.5.2. Sparse LDA**

L'analyse discriminante linéaire classique (LDA) ou LDA de Fisher, aborde le problème de classification en cherchant des projections mappant un vecteur de données de grande dimension, dans le sous-espace de faible dimension le plus discriminant. L'analyse discriminante Sparse effectue une analyse discriminante linéaire avec un critère de rareté imposé de telle sorte que la classification et la sélection des caractéristiques soient effectuées simultanément. Elle est basée sur l'interprétation de score optimale de l'analyse discriminante linéaire. Elle peut être étendue pour effectuer une discrimination clairsemée (*sparse*) via des mélanges de gaussiennes si les limites entre les classes ne sont pas linéaires ou si des sous-groupes sont présents au sein de chaque classe [27] [28].

## **1.5. Conclusion**

L'un des principaux buts de la réduction de dimension est de faciliter la visualisation et la compréhension des données. De plus, la réduction de dimension consiste à la fois, à extraire et sélectionner un sous-ensemble optimal de caractéristiques pertinentes et d'éliminer les informations redondantes pour un critère fixé auparavant. Cette sélection ou extraction conduit donc à la réduction de l'espace de stockage nécessaire et le temps d'apprentissage et d'utilisation afin de rendre l'ensemble de données plus représentatif ainsi que facile à traiter [29].

Dans ce chapitre on a introduit la réduction de dimension et ses différentes techniques, notamment l'analyse discriminante linéaire, ses algorithmes, et ses variantes. L'ADL a été exposée comme étant elle réduit la dimensionnalité d'un ensemble de données, ainsi que pour la reconnaissance faciale. Celle-ci est employée pour appliquer l'ADL, dans ce présent travail.

# *CHAPITRE 2*

## *Identification de visages*

### **2.1. Introduction**

La reconnaissance de visage est un domaine d'intérêt majeur à nos jours. Elle est définie comme étant le processus biométrique appliqué sur des images ou vidéo. Il existe de nombreux algorithmes de reconnaissance de visage qui, généralement extraient des caractéristiques et les comparent à une base de données pour estimer la meilleur correspondance [30]. Ces algorithmes sont utilisés dans un large éventail d'applications comme: contrôle de sécurité, enquête criminelle et contrôle d'entrée dans les bâtiments ou d'accès aux guichets automatiques, vérification des passeports, identification des visages d'une base de données donnée... etc. [31].

Dans ce chapitre nous présentons de brèves définitions concernant la biométrie et ses types. Par la suite, nous exposons les différents types de reconnaissance faciale notamment l'identification de visage par l'algorithme de technique ADL.

### **2.2. Biométrie et reconnaissance de visage**

La biométrie est l'utilisation automatisée de caractéristiques physiologiques ou comportementales pour déterminer ou vérifier ou identifier l'identité d'une personne [32]. Les systèmes biométriques offrent plusieurs avantages par rapport aux mesures de sécurité traditionnelles. L'apparition de l'ordinateur et sa capacité à traiter et à stocker les données, a largement permis la création des systèmes biométriques informatisés [33]. Il existe plusieurs technologies biométriques comportementales, on en cite :

- \* reconnaissance vocale
- \* dynamique de frappe au clavier
- \* démarche
- \* reconnaissance de gestes

\* dynamique du tracé de la signature (vitesse de déplacement du stylo, accélérations, pression exercée, inclinaison)...etc.

Cependant, les mesures physiologiques se divisent en mesures biologiques ou morphologiques. Les analyses biologiques les plus répandues sont l'ADN, le sang, la salive, ou l'urine dans le domaine médical, pour des investigations criminelles ou dans le domaine du sport pour des contrôles de dopage [34]. Par contre, les mesures morphologiques regroupent les empreintes digitales, la forme de la main, du doigt, le réseau veineux, l'iris, la rétine et le visage [33]. Ce dernier constitue l'objectif de notre expérience. Étant donné ses multiples avantages, notamment sa précision qui permet des taux d'inscription et de vérification élevés, le visage est également la caractéristique physiologique la plus naturelle pour se reconnaître. Toutefois, comme les visages peuvent être capturés à une certaine distance, la reconnaissance faciale peut être effectuée sans aucun contact physique et ne nécessite aucune interaction physique de la part de l'utilisateur [32].

Les différentes technologies utilisées font l'objet de recherches régulières, de développements et bien entendu, d'améliorations constantes. Pourtant, les différents types de mesures n'ont pas le même niveau de fiabilité. D'ailleurs, il est estimé que les mesures physiologiques ont l'avantage d'être les plus stables dans la vie d'un individu [34].

### **2.3. Reconnaissance de visages**

Le visage est une structure multidimensionnelle complexe, il nécessite une technique informatique intelligente pour la reconnaissance. Lors de l'utilisation d'un système automatique de reconnaissance faciale, la position du visage par rapport à la caméra, les changements d'éclairage, et les variations de poses sont cruciales pour la détermination de l'identité de l'individu [30]. En général, les systèmes biométriques de reconnaissance faciale fonctionnent en deux modes : Vérification et Identification [33]. Il existe une différence subtile entre le mode d'identification de visage et ce de vérification :

\*\* l'identification s'introduit lorsqu'un visage inconnu est présenté au système, et il est prévu de rapporter l'identité de l'individu à partir d'une base de données de visages.

\*\*Alors que dans la vérification, il y a une identité revendiquée soumise au système, qui doit être confirmée ou rejetée [36].

Le schéma fonctionnel général d'un système de reconnaissance faciale est composé de quatre processus : Le visage est d'abord détecté de la donnée 2D. Puis le visage extrait est aligné (par normalisation de la taille). Les caractéristiques discriminantes sont ensuite extraites afin d'être mises en correspondance avec les utilisateurs inscrits dans la base de données du système, la sortie du système est le visage ID de l'image de la personne donnée [36].

### 2.3.1. Différentes phases d'un système de reconnaissance de visages

1. La première phase de tout système reconnaissance de visages est la phase d'enrôlement, c'est la phase d'acquisition de données. Pendant laquelle, un utilisateur est enregistré dans le système pour la première fois. Elle sert notamment à acquérir un modèle d'identification. Ce dernier est stocké et utilisé dans la phase d'identification pour une comparaison avec les informations entrées au système. Ce stockage de modèle peut s'accompagner par l'ajout d'information biographique dans la base de données [37]. Après le recueil des données de base, le système garde les données du prétraitement dans une base de modèles. Ces données seront utilisées en comparaison afin d'identifier une entité [35]. La figure 2.1 illustre une procédure de reconnaissance faciale typique.

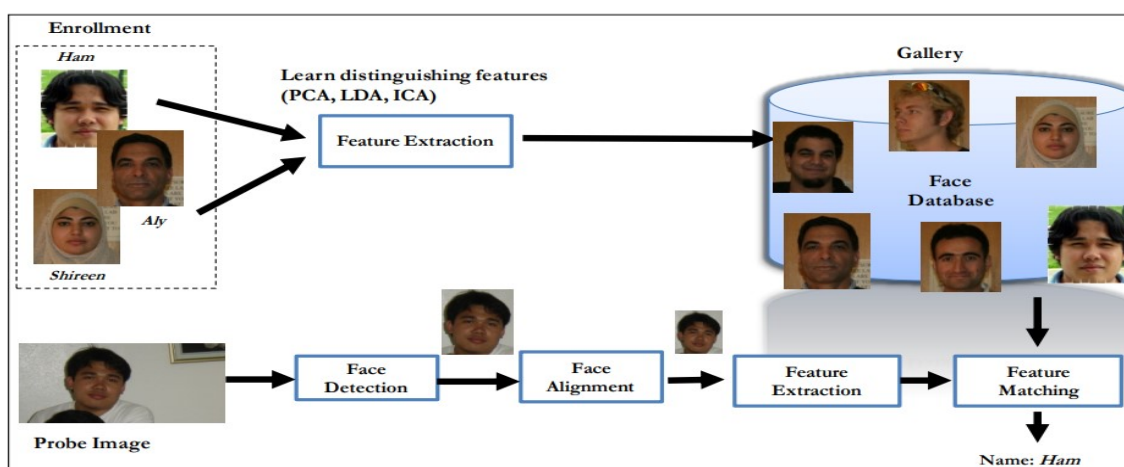


Figure 2.1 : Processus de reconnaissance faciale [36].

2. Lors de l'identification, le modèle recueilli n'est plus stocké mais comparé aux modèles déjà existants, afin de d'identifier une entité. A la fin de cette comparaison une décision est prise et l'entité est, soit rejetée, soit confirmée.

Lorsqu'un système biométrique opère en mode authentification ou vérification, l'utilisateur affirme son identité et le système vérifie si cette affirmation est vraie (client) ou fausse (imposteur).

Les applications de systèmes de reconnaissance de visage sont nombreuses, soit dans le secteur public ou le secteur privé [38]. Les applications dans le secteur public sont principalement des applications mises en place par les autorités nationales notamment : Les applications de loi et sécurité nationale, Sécurité à la frontière, Applications militaires ... etc. [34].

Concernant le secteur privé, nous citons : Sécurité des appareils mobiles, Services en ligne, Sécurité chez soi ...etc. [38]. Le Smartphone étant un appareil privé et individuel, Android dans sa quatrième version a intégré l'authentification par reconnaissance de visage. Cela permet à l'utilisateur de déverrouiller son

téléphone en le mettant face à son visage. *Iphone* également a intégré dans toutes ses nouvelles versions, une technologie de reconnaissance de visage de haute précision nommée *Face ID9* [35].

### 2.3.2. Différentes méthodes de reconnaissance de visages

Le visage, en tant que partie du corps humain, est le plus facile et le plus souvent utilisé pour distinguer l'identité des individus [39]. Nous pouvons reconnaître les visages avec la vision par ordinateur en utilisant une variété de modèles et d'algorithmes [30].

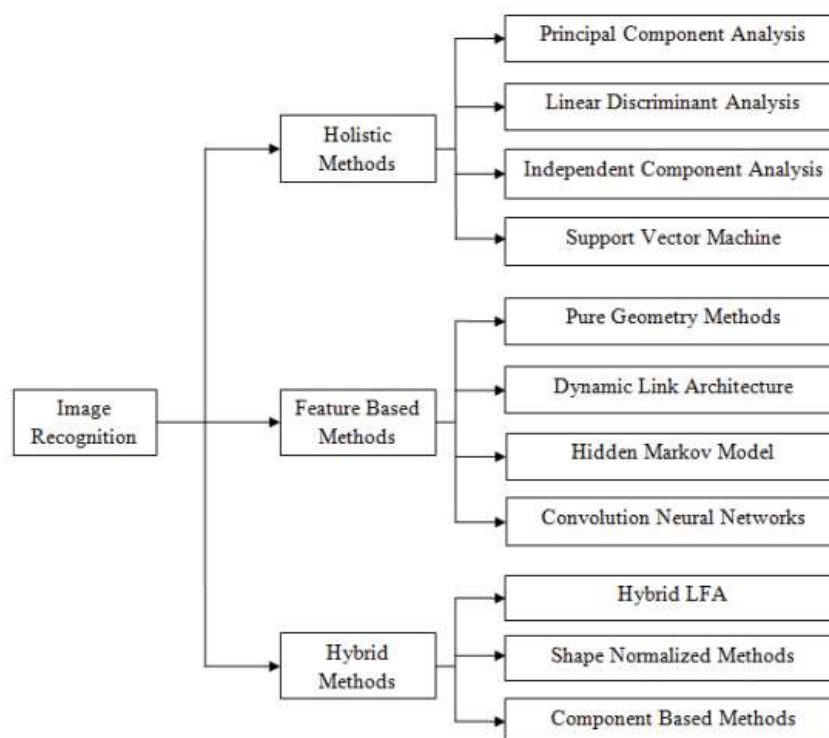


Figure 2.2 : différentes méthodes déployées dans un système de reconnaissance faciale [31].

Généralement, le système de reconnaissance faciale est divisé en 3 méthodes illustrées par La Figure 2.2 ci dessous: méthodes holistiques (ou globales), méthodes basées sur les caractéristiques dites méthodes locales, méthodes hybrides, [31].

- \* Les méthodes holistiques utilisent toute la région du visage comme entrée du système de reconnaissance [36].
- \* Dans les méthodes basées sur les caractéristiques, les caractéristiques locales telles que les yeux, le nez et la bouche sont d'abord extraites. Ensuite, leurs emplacements et statistiques locales (géométriques et/ou d'apparence) sont introduits dans un classificateur structurel.

\* La dernière catégorie des méthodes, tout comme le système de perception humaine, les méthodes hybrides utilisent à la fois des caractéristiques locales et toute la région du visage pour reconnaître un visage [31].

### 2.3.3. Identification de visage

Un système d'identification faciale doit être capable d'identifier automatiquement les visages dans les images ou les vidéos. Il donne l'identité d'une personne en correspondant ses propres données (échantillon), avec celles recueillies dans une base de données lors de l'enrôlement [35]. Cependant, l'utilisateur ne dévoile pas son identité. L'affirmation faite par l'utilisateur est qu'elle est une des personnes déjà enrôlées par le système. Ainsi, le visage de l'individu est comparé avec les modèles de toutes les personnes de la base de données. On parle alors de correspondance 1 à N. Si l'échantillon correspond à celui d'une personne de la base, la personne est identifiée.

Généralement, on peut évaluer les performances d'un système d'identification en calculant le taux de reconnaissance du système défini par [40]:

$$\text{Taux de reconnaissance} = \frac{\text{Nombre de décisions correctes}}{\text{Nombre total de décisions}}$$

### 2.3.4. Courbe CMC

La courbe CMC pour *Cumulative match characteristic*, est la courbe de correspondance cumulative, C'est un indice d'évaluation important pour les systèmes de reconnaissance de formes, tels que les visages, les empreintes digitales et l'iris, en particulier dans le système d'identification biométrique [41]. Cette courbe présente les valeurs du rang d'identification et les accumulations des décisions correctes, respectivement en abscisses et en ordonnées. Elle est utilisée pour comparer la performance de systèmes d'identification biométrique [42]. Un exemple de cette courbe est donné par la figure suivante.

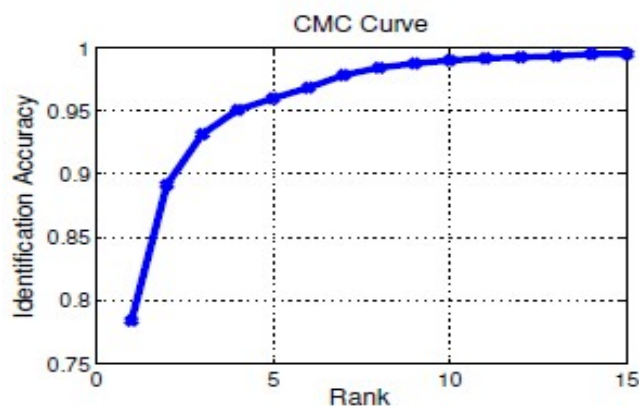


Figure 2.3 : Exemple de la courbe CMC [43].

### 2.3.5. Algorithme d'identification par ADL

L'algorithme d'identification par ADL est divisé en deux phases, l'une est utilisée pour calculer le modèle des individus concernés, appelée phase d'apprentissage du système, et l'autre comprend l'identification des individus test par rapport aux modèles enregistrés à l'entrée du système, que nous appellerons étape de test.

Les étapes de classification des identités sont les suivantes [39] :

- Conversion de l'image du visage testée par la taille de  $N * N$  en vecteur colonne de forme  $r_{i N^2}$
- Normalisation de l'image du visage d'entrée à l'image d'entraînement en soustrayant la valeur moyenne de l'image d'entraînement. pour trouver la valeur des différentes matrices  $\Phi_{inp}$ .
- Calcul le poids de l'image test en multipliant la matrice transposée des valeurs propres  $\vec{V}^T$  par la matrice  $\Phi_{inp}$  matrice ( $U_{inp}$ ) ;  $U_{inp} = \vec{V}^T \Phi_{inp}$
- Calcul de la distance de la différence entre l'image testée avec l'image du visage d'apprentissage en utilisant par exemple la distance euclidienne.

$$\varepsilon_i = \sqrt{\|U - U_{inp}\|^2}; i = 1 \dots M$$

- Le résultat de l'identification est l'image qui a la plus petite distance avec l'image de test affichée par le système.

Pour déterminer si le système fonctionne bien et correctement, il est nécessaire de tester les processus suivants :

#### 1. Processus d'apprentissage

C'est la première étape du test du système. Cette étape vise à générer la valeur pondérale de chaque image d'apprentissage.

#### 2. Processus d'identification de visages

Une fois le processus d'apprentissage est achevé avec succès, l'étape suivante consiste à effectuer le processus de d'identification de visages. L'objectif est de déterminer la capacité du système à reconnaître avec succès l'image de test ou à tester correctement. On cite deux cas :

- *L'image d'entraînement est la même que l'image de test*

L'image de test utilisée par le système dans le processus d'identification est la même que l'image d'apprentissage utilisée au moment du processus d'apprentissage, à la fois en termes d'expression, de position et de condition d'image.

- L'image d'entraînement n'est pas la même que l'image de test.

Dans ce cas, l'image de test utilisée n'est pas la même que l'image d'apprentissage utilisée lors de l'apprentissage, qu'elle soit vue en termes d'expression, de position et d'état de l'image. Mais l'image de test et l'image d'apprentissage proviennent toujours de l'image du visage de la même personne.

### 2.4. Etat de l'art de l'identification de visages par ADL

Dans la l'identification faciale informatisée, chaque visage est représenté par un grand nombre de valeurs de pixels. L'analyse discriminante linéaire est principalement utilisée ici, pour réduire le nombre de caractéristiques à un nombre plus gérable avant la classification. Chacune des nouvelles dimensions est une combinaison linéaire de valeurs de pixels, qui forment un modèle [31].

L'algorithme ADL est appliqué aux images en 1997 par Belhumer et al à Yale University aux USA, aussi connu sous le nom de Fisherfaces, il permet d'effectuer une véritable séparation de classes, contrairement à l'ACP [44]. La figure suivante décrit le processus d'identification de visage par ADL :

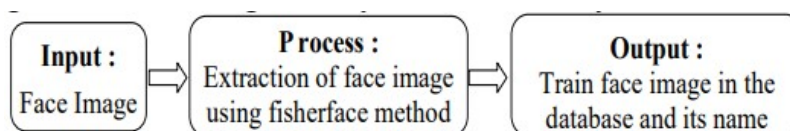


Figure2.4 : Étapes du processus d'identification faciale par ADL [39].

L'ADL est étroitement lié à l'ACP, du fait que toutes les deux recherchent les combinaisons linéaires des variables qui représentant au mieux les données. Elle essaye explicitement de modéliser la différence entre les classes de données contrairement à l'ACP qui ne tient pas compte des différences entre les classes.

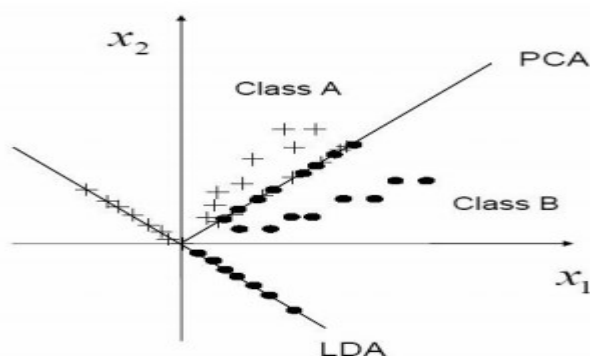


Figure 2.5 : Différence entre le classement de données par ADL, et PCA [45]

L'ADL nous fournit un petit ensemble de caractéristiques qui contiennent les informations les plus pertinentes à des fins de classification. Les caractéristiques sont obtenues par analyse de vecteurs propres de matrices de dispersion et le résultat est un schéma efficace d'extraction et de classification de caractéristiques

basé sur des projections [46]. Cependant l'algorithme d'ADL ne fournit pas des solutions fiables et robustes au problème de linéarité souvent rencontré en reconnaissance faciale, Plusieurs approches basée sur l'ADL ont été proposées, ces méthodes ont prouvé leur fiabilité et ont permis largement la résolution des problèmes d'ADL. Les algorithmes basés sur KDA (l'extension non linéaire de LDA avec l'astuce du noyau *Kernel Discriminant Analysis*) et KDDA (l'analyse discriminante directe du noyau *Kernel Direct Discriminant Analysis*) traitent la non-linéarité de la distribution des motifs du visage et résolvent efficacement le problème de « petite taille d'échantillon » (SSS), l'algorithme KDDA a été proposé dans [47] pour la reconnaissance faciale, l'algorithme généralise les points forts d'ADL directe (D-LDA) et des techniques du noyau tout en surmontant bon nombre de leurs défauts et limitations, en suivant le paradigme SVM, l'espace d'entrée d'origine est mappé d'abord de manière non linéaire à un espace de caractéristiques implicite de haute dimension, où la distribution des motifs de visage devrait être linéarisée et simplifiée. Ensuite, une nouvelle variante de la méthode D-LDA est introduite pour résoudre efficacement le problème SSS et dériver un ensemble de vecteurs de base discriminants optimaux dans l'espace des caractéristiques [47]. Dans KDA les chercheurs utilisent l'astuce du noyau pour mapper les données d'entrée à un espace de noyau de dimension supérieure afin de représenter une relation non linéaire complexe des données d'entrée. Ensuite, KDA trouve une projection linéaire optimale de l'espace du noyau vers un sous-espace de projection de dimension inférieure dans lequel toutes les classes sont bien séparées des autres. Certains algorithmes ont été proposés pour améliorer les performances de KDA [53]. La KODA (Analyse discriminante basée sur l'optimisation du noyau *Kernel Optimization-based Discriminant Analysis*) a été proposée dans [53] pour améliorer KDA en reconnaissance faciale. La procédure de KODA est divisée en deux étapes : l'optimisation du noyau et la projection. La KODA ajuste automatiquement les paramètres du noyau en fonction des échantillons d'entrée et les performances d'extraction de caractéristiques sont améliorées pour la reconnaissance faciale. Des simulations sur les bases de données de visages Yale et ORL ont démontrées la faisabilité d'améliorer KDA avec l'optimisation du noyau. Pour plus de détails voir [53].

## **2.5. Conclusion**

Actuellement, la reconnaissance de visage est la technologie biométrique la plus déployée sur le marché. Elle permet un déploiement rapide et à grande échelle grâce à sa facilité de mise en œuvre, tout en réduisant les coûts. Cette technologie a permis également une grande précision pour l'identification des personnes et une grande rapidité d'exécution [40].

On a abordé dans ce chapitre, quelques notions de la biométrie. Par la suite, on a cité les différentes techniques d'identification faciale, notamment, l'analyse discriminante linéaire.

# *CHAPITRE 3*

## *Simulations et discussions*

### **3.1. Introduction**

Les méthodes de réduction de dimension linéaires sont le socle de l'analyse de données de grande dimension, en raison de leurs interprétations géométriques simples et de leurs propriétés de calcul généralement attrayantes. Néanmoins, la représentation des caractéristiques de faible dimension avec un pouvoir de discrimination est primordiale dans les systèmes d'identification faciale [45].

La représentation de visages par ADL est une approche de type image. Chaque image de visage est considérée comme un vecteur dans un espace ayant autant de dimensions que de pixels dans l'image. Les caractéristiques de l'image sont extraites par cette méthode mathématique de réduction de dimensionnalité.

Dans ce chapitre, nous allons appliquer l'ADL, dans le but de réduire la dimension de l'espace des visages, d'une part. D'autre part, on veut établir une identification faciale en utilisant la base de données ORL.

### **3.2. Base de données ORL**

La base de données *Olivetti Research Laboratory* (ORL) des images visages, est constituée d'un ensemble d'images de visages prises entre avril 1992 et avril 1994 au laboratoire de recherche Olivetti Research Laboratory à Cambridge, UK [48]. La base contient 10 images de 40 personnes distinctes. Pour certaines personnes, les images ont été prises à des moments différents, en variant l'éclairage, les expressions faciales (yeux ouverts/fermés, souriant/non souriant) et les détails du visage (lunettes/pas de lunettes).

La Figure.3.1, illustre la totalité des images cette base. Toutes les images ont été prises sur un fond sombre et homogène avec les sujets en position verticale et frontale (avec une tolérance pour certains mouvements latéraux) [49].

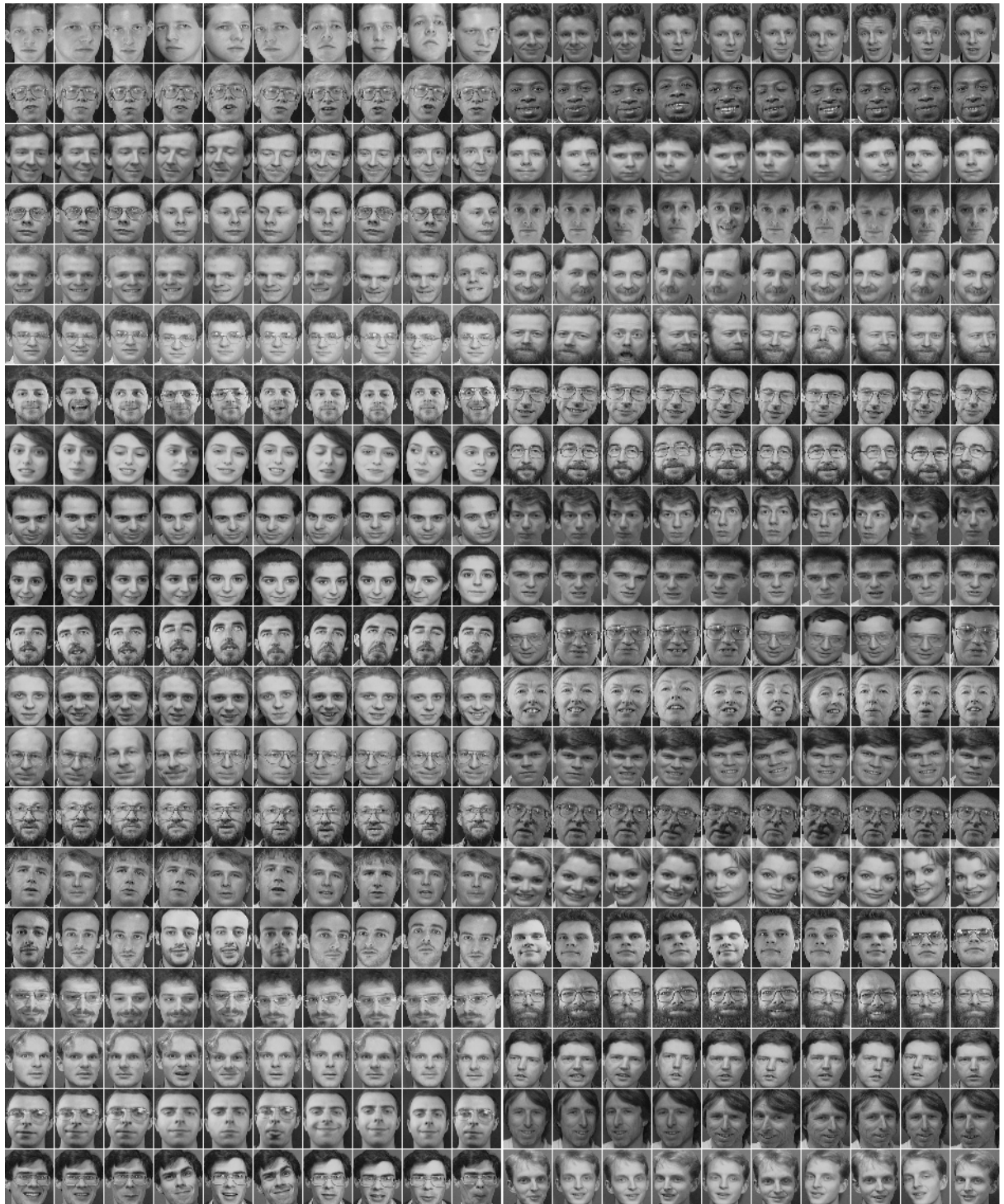


Figure 3.1 : Total des images contenues dans la base de données ORL [49].

Les images sont organisées en 40 répertoires (un pour chaque sujet. Chaque répertoire contient dix images différentes de ce sujet, qui ont des noms de la forme Y.pgm, où Y est le numéro d'image pour ce sujet (entre 1 et 10) [48]. Les statistiques de l'ensemble de données ORL [50] [51] sont:

- \* 400 images,
  - \* 40 sujets,
  - \* Couleur : niveaux de gris
  - \* Taille de l'échantillon (image): 92 x 112 pixels
  - \* Variations d'éclairage, occlusions, expression faciale.
- \* Taille de l'ensemble de données : 4,5 Mo (compressé sous forme de fichier ZIP)



Figure.3.2. Images des dix premières personnes de la base ORL



Figure.3.3 : Exemple de changements d'éclairage

La base de données ORL prend aussi en considération les expressions faciales, telles que les grimaces. En voici un exemple illustrée par la Figure.3.4. :



Figure.3.4 : Exemple de changements des expressions faciales

Voici un exemple où l'acquisition se fait sous différentes orientations du visage et sous différents éclairages :



Figure.3.5 : Exemple de changements d'orientations du visage

L'exemple qui suit, exemple montre les changements d'échelle dus à la distance entre le dispositif d'acquisition et l'individu :



Figure.3.6 : Exemple de changements d'échelle

### 3.3. Algorithme et programmation

Dans cette partie, on va présenter l'algorithme utilisé dans nos simulations ainsi que le programme dédié à cet algorithme. Les étapes suivies pour appliquer la réduction de dimensionnalité, sont :

1. Lecture des images et chargements des données.
2. Extraction de données et réduction par l'ADL.
3. Identification des visages (tests et comparaison) par mesure de distance.
4. Affichage des résultats.

En ce qui concerne la programmation, on s'est basé sur le programme PhD face recognition toolbox . Celui-ci est une boîte à outils de reconnaissance faciale ; PhD pour (Pretty Helpful Development functions for). Elle est constituée d'une collection de fonctions et de scripts Matlab destinés à aider les chercheurs travaillant dans le domaine de la reconnaissance faciale [52].

La phase de programmation est, généralement, divisée en 6 parties. La première partie est consacrée à la lecture et le stockage de données (images), sous forme de fichiers.mat. Ces derniers sont par la suite chargés par le programme principal.

La seconde partie, établit la procédure ADL et crée le modèle ADL pour la réduction et la classification. Ce modèle change avec le choix de l'ensemble d'images d'apprentissage donné par la matrice de données  $X$ , de taille  $B \times A$ . Chacune des colonnes  $A$ , contient un échantillon (visage) ayant une dimensionnalité de  $B$ .

En sortie, le modèle ADL délivre les paramètres du sous-espace ADL nécessaires pour projeter un échantillon de test donné dans ce dernier. Le modèle a la forme d'une structure regroupant les paramètres suivants :

- \* *model.P* : le visage moyen sous forme vectorielle.
- \* *model.dim* : la dimensionnalité du sous-espace ADL
- \* *model.W* : la matrice de transformation (les visages ADL)
- \* *model.train* : caractéristiques ADL correspondant aux données d'entraînement

Ce programme applique en premier lieu l'ACP comme étape de prétraitement, afin d'éviter le problème de petite taille d'échantillons qui se résulte ici du fait que le nombre de dimension de l'ensemble de données est très élevé par rapport à la taille de différents ensembles d'apprentissage (nombre d'échantillons *training*) utilisés.

La troisième partie, consiste à calculer la distance entre les différentes images d'apprentissage et images de test, afin de choisir en fonction de cette distance l'image appropriée. Autrement dit, plus la distance entre les images est petite plus l'identification est meilleure. L'image la plus proche révèle l'identité de l'image test.

Les résultats finaux de l'identification illustrant la performance du programme, sont affichés au moyen de la courbe CMC.

### 3.4. Simulations et discussions

#### 3.4.1. Aspect matériel et logiciel

Ce présent travail a été développé sur un PC de Processeur : Intel(R) Celeron (R) CPU N3350 @ 1.10GHz. Sa capacité Mémoire (RAM) :4.00 Go. Sa Capacité disque dur : 500 Go. Il est doté d'un Système d'exploitation : 64 bits, processeur x64 sous Windows10 Professionnel version 20H2.

L'environnement de développement utilisé est celui de MATLAB Version 9.2 (R2017a). Cette plateforme de calcul numérique et de programmation est utilisée par des millions d'ingénieurs et de scientifiques pour analyser des données, développer des algorithmes et créer des modèles. Il est conçu pour l'analyse par itération et les processus de conception avec un langage de programmation qui permet d'exprimer directement les mathématiques sous forme de tableaux et de matrices. La Figure 3.7, donne un exemple d'exploitation de cet environnement.

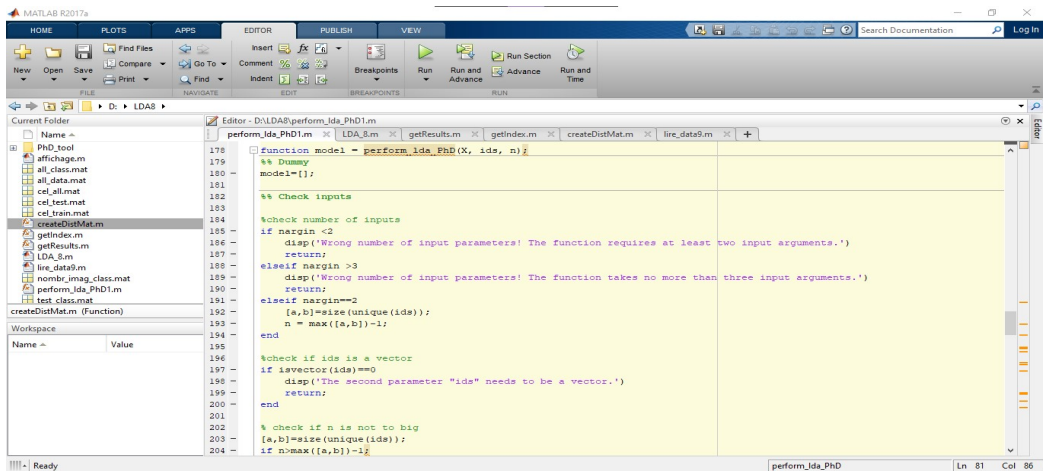


Figure.3.7 : Aperçus sur la fonction *perform\_lda\_PhD* utilisée pour la création du model ADL

Parmi les raisons qui nous ont poussés à l'utiliser, on trouve entre autres: ses très nombreuses fonctions prédéfinies et prêtes à l'emploi, sa simplicité à l'implémentation et rapidité de calculs, sa fiabilité et sa robustesse [33].

### 3.4.2. Simulations

Pour vérifier l'efficacité de notre application, on a effectué 3 principaux tests, dans les deux derniers tests, l'expression faciale et les détails du visage, ont fait l'objet principal du choix d'images de la base d'apprentissage, ainsi que celles consacrée à l'ensemble test.

#### Remarques importantes

1. Pendant, les premières exécutions effectuées dans le but de tester le programme final, le problème d'insuffisance mémoire (*Out of memory*) s'y présenté. Celui est aux caractéristiques matérielles disponibles (4G de RAM). Ce problème persistait même en utilisant 5 classes seulement. Cependant, pour contourner cette limitation, on a redimensionné les images à la taille (61x51). Un petit exemple est illustré par la figure suivante. Ce redimensionnement est accompagné par la perte de certaines caractéristiques discriminatoires.

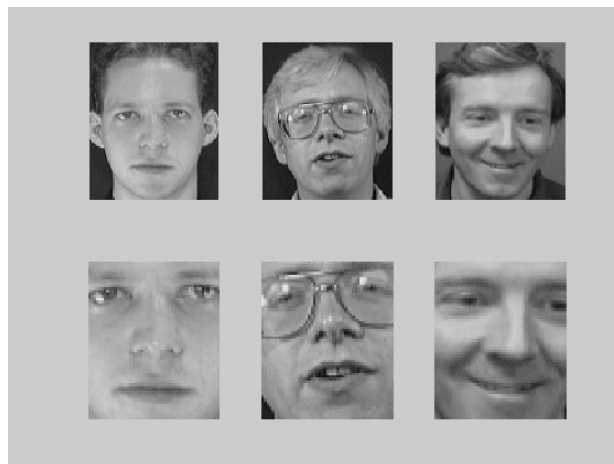


Figure 3.8: Exemple de 3 images redimensionnées.

2. Dans tous les tests qui vont suivre, le nombre de vecteurs propres (prétraitement par ACP) conservés est égal au nombre d'échantillons constituant l'ensemble training. De même, le nombre de caractéristiques discriminantes retenues par ADL est égal au nombre de classes distinctes utilisées en apprentissage moins 1. Cela implique, si par exemple, on utilise 200 échantillons (61x51) en training appartenant aux 40 classes, l'espace réduit par ADL est de dimension 39. rappelons que l'espace de départ a une dimension égale à  $61 \times 51 = 711$ .

### 3.4.2.1. Ensemble d'apprentissage 1 (200 : 200)

Dans ce premier test, les 5 premières images des 40 individus constituent l'ensemble l'apprentissage avec un total de 200 échantillons. Par contre les 5 dernières images (200 images), sont utilisées pour le test. Le résultat de l'identification est illustré par la courbe CMC de la Figure 3.9 ci après.

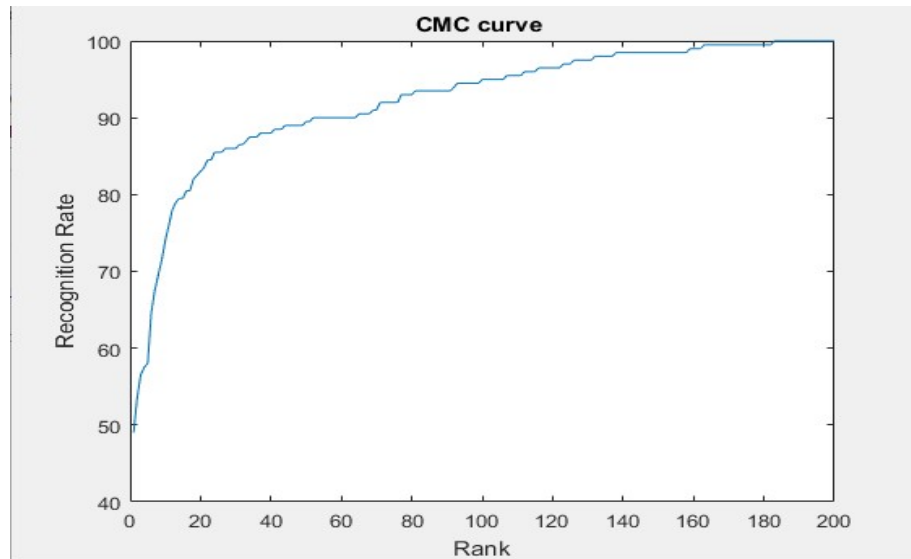


Figure 3.9: Courbe CMC pour l'ensemble d'apprentissage 1.

### 3.4.2.2. Ensemble d'apprentissage 2 (200 : 200)

Dans ce test, on n'a permuté les ensembles d'apprentissage et de test de l'essai précédent. Cependant, les 5 premières images des 40 individus constituent l'ensemble de test. Tandis que, 5 dernières images (200 images), sont utilisées pour le training. La courbe CMC correspondante est illustrée par la Figure 3.12, ci après.

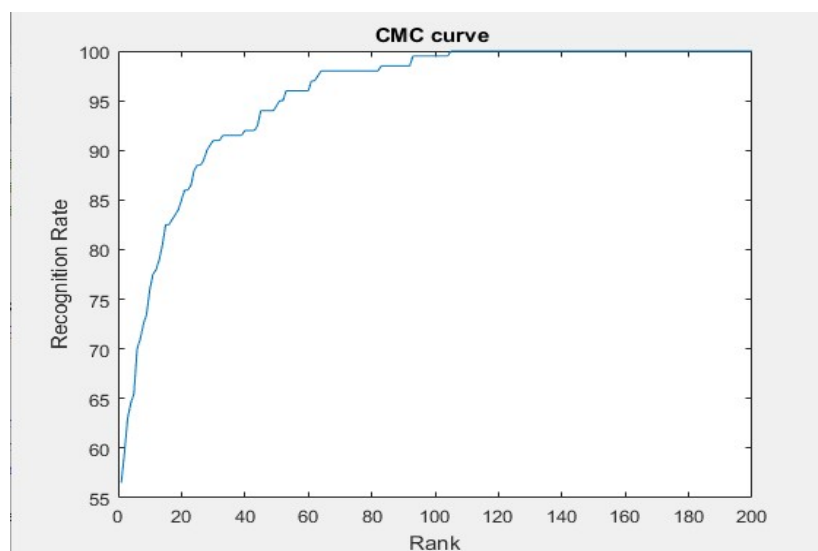


Figure 3.10: Courbe CMC du test réciproque

### 3.4.2.3. Ensemble d'apprentissage 3 (280 : 120)

Dans ce cas, l'ensemble d'apprentissage est composé de 280 images ; notamment 3 images pour chaque personne. Les 3 images restantes pour chacun sont regroupées dans l'ensemble test. Le résultat de cette simulation est présenté par la Figure 3.13 ci après.

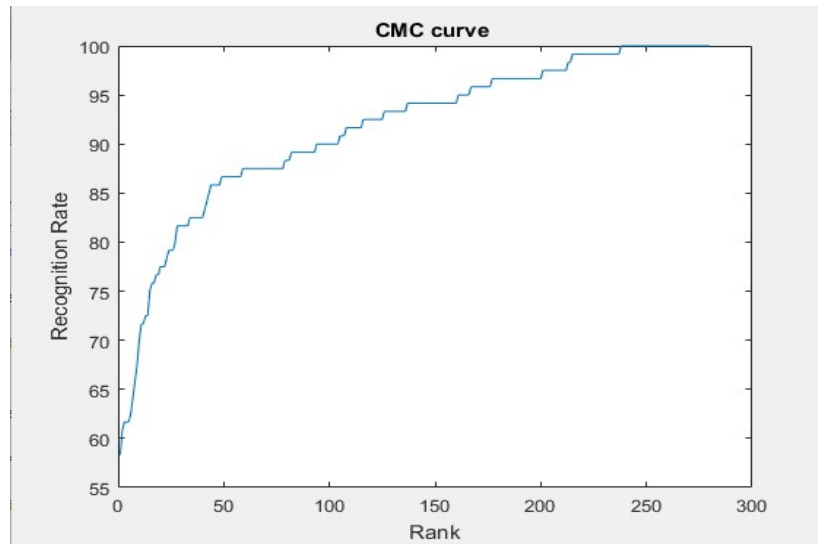


Figure 3.11: Courbe CMC pour l'ensemble d'apprentissage 3

### 3.4.2.4. Ensemble d'apprentissage 4 (360 : 40)

Actuellement, On prend les 9 premières images comme ensemble d'apprentissage. La dernière image de chaque individu est utilisée pour le test. La courbe CMC de la Figure 3.14 affiche le taux de l'identification.

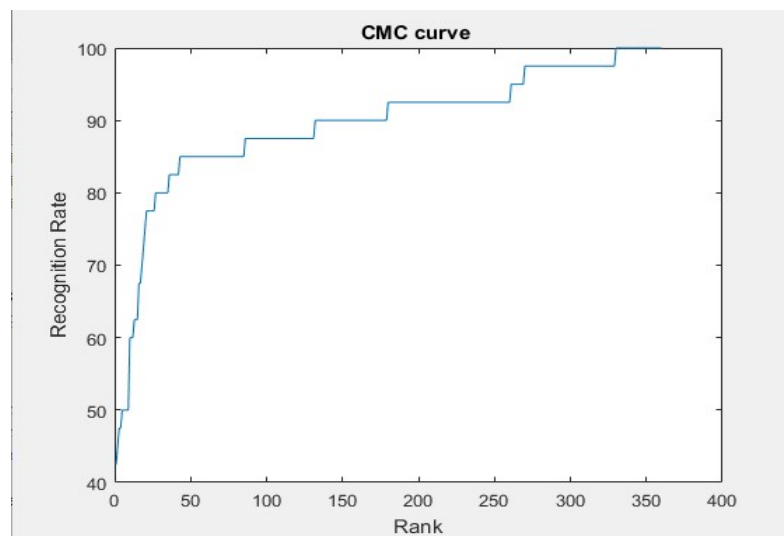


Figure 3.12 : Courbe CMC pour l'ensemble d'apprentissage 4

### **3.4.3. Discussions des résultats**

1. Comme nous avons déjà mentionné dans le premier chapitre, notre ADL souffre du problème SSS car la taille des images utilisées est beaucoup plus grande que nombre d'échantillons pris dans tous les tests précédents. Les tests effectués sans ACP (non reportés dans ce manuscrit), n'ont pas pu aboutir à la représentation de l'espace ADL (*null space problem*).

2. En examinant les différentes mesures de performance illustrées par les courbes CMC, On peut révéler ce qui suit:

- \* les taux de d'identifications, sont jugés assez bas. Cela est du à plusieurs facteurs. L'un des facteurs majeurs est la réduction considérable de l'espace ; à noter la conservation de 39 vecteurs ADL seulement. Cela implique la négligence de certaines informations primordiales du point de vu discriminante.

- \* le redimensionnement des images a diminué la qualité séparable des caractéristiques représentatives des classes. Cela influe sur la qualité de la matrice inter-classes.

- \* la base de données utilisée, présente des variances importantes dues à l'éclairage, les expressions, l'échelle... etc, au sein de la même classe. Cela est projeté directement sur la matrice intra-classes, dans laquelle on souhaite minimiser les discriminations le plus que possible. Le choix des ensembles d'apprentissage est fait sans analyse préalable de l'influence de des groupes testés sur matrice intra-classes. Il a fallu établir plus de combinaisons tout en considérant la variabilité inter et intra-classes .

- \* l'utilisation de l'ACP comme prétraitement est issue d'un compromis délicat. D'une part, il faut éviter le problème SSS dont les valeurs propres significantes sont prises en considération. Notons que les valeurs propres négligées peuvent être importantes du point de vue séparabilité (critères d'optimisation pour ACP et ADL différents). D'autre part, il oblige l'ADL à manipulé des données propres se basant sur la caractéristiques corrélation et non pas séparabilité.

En fin, le choix du nombre d'images utilisées en apprentissage et en test, la position, ainsi que les expressions et détails des visages, représentent le critère principal d'évaluation de performance d'identification dans un système de reconnaissance faciale.

### **3.5. Conclusion**

Ce chapitre, a mets en évidence le processus de réduction de dimensions de données faciales, à l'aide de l'analyse discriminante linéaire qui, comme toute autre méthode de réduction de dimensionnalité, a ses avantages et ses inconvénients qu'l'on a essayé d'illustrer dans nos test et simulations.

Dans ce chapitre, la technique de l'analyse discriminante linéaire ADL a été appliquée dans le but de représenter des images de la base ORL dans un espace de dimension réduite. Quatre variantes ont été

exposées suivant l'ensemble d'images d'apprentissage utilisées. Pour chacune, des mesures sont effectuées. Comme toute autre méthode de réduction de dimensionnalité, l'ADL a ses avantages et ses inconvénients qu'il on a essayé d'illustrer dans nos test et simulations.

*Conclusion*

*Générale*

## *Conclusions générale*

Dans la reconnaissance faciale informatisée, chaque visage est représenté par un grand nombre de valeurs de pixels. L'analyse discriminante linéaire ADL, est principalement utilisée ici pour accentuer la séparabilité entre les différentes classes, tout en augmentant la ressemblance au sein de chaque classe. Aussi, elle vise à réduire le nombre de caractéristiques à un nombre plus gérable avant la classification. Chacune des nouvelles dimensions est une combinaison linéaire de valeurs de pixels, qui forment un modèle.

Nous avons présenté un système répondant à l'objectif que nous nous sommes fixés au départ, à savoir un système permettant la réduction de dimension appliqué à l'identification d'individus, nous avons également souligné les avantages et les inconvénients de la technique exercée dans ce travail en se basant sur les tests et résultats qu'on a obtenus durant la phase de programmation et simulation.

Nous estimons que la réduction de dimension est un domaine d'intérêt majeur ces jours-ci, surtout dans les applications d'exploration de données à grandes dimensions. Les algorithmes des méthodes de réduction de dimensionnalité sont utilisés dans un large éventail d'applications telles que les signaux vocaux, les scans IRMf, l'analyse des signaux d'électrocardiogramme, la récupération multimédia, les applications basées sur le marché, ...etc.

Nous avons souligné durant ce travail l'influence néfaste du changement, d'expressions de visages et d'éclairage sur le taux de reconnaissance du système. Il existe plusieurs méthodes qui sont étudiées et développées par plusieurs spécialistes dans ce domaine, en guise d'améliorer la performance des systèmes classiques. Une méthode qui a suscité les efforts, c'est le multimodale, dans lequel on combine plusieurs technologies biométriques ou plusieurs algorithmes de reconnaissance.

Une première extension est d'améliorer la procédure ADL en combinant quelque unes de ces variantes tout en adaptant les choix à l'application considérée. Deuxièmes, ce travail peut être réalisée en appliquant la technique employée dans ce mémoire, dans un système de sécurité biométrique utilisant l'acquisition vidéo. Dans ce cas, cet algorithme doit répondre aux exigences du temps réel, et cela ne peut se réaliser sans le développement d'algorithmes efficaces de détection et de poursuite des visages dans les séquences vidéo.

- [1]. Quanquan Gu, Zhenhui Li, and Jiawei Han, « *Linear Discriminant Dimensionality Reduction* », University of Illinois at Urbana-Champaign, Urbana, IL 61801, 16 p.
- [2]. Yann Teyssier, « *Intérêt de la réduction de dimension pour évaluation de la fibrose hépatique en échographie : approche radiomique basée sur les contours du foie* », thèse pour l'obtention du titre de Docteur en médecine, Médecine humaine et pathologie, Université Grenoble Alpes, Médecine humaine et pathologie, 2019. dumas-02954552, pp.90.
- [3]. BELAHCEN, M, *Méthodes de réduction et classification*, dans, « Authentification et identification en Biométrie », thèse de doctorat en sciences en Automatique, sous la direction de Deibilou AbdErrazak, Biskra, université MohamedKhider, 2013, pp.28.
- [4] Omprakash Saini, Prof. Sumit Sharma, « *A Review on Dimension Reduction Techniques in Data Mining* », Computer Engineering and Intelligent Systems, Vol.9, No.1, 2018, pp.8. En ligne:
- [5]. Olivier, Godin, « *Réduction de la dimension des données* », Université de Sherbrooke, Sherbrooke, 2020, pp.109.
- [6]Olivier, Godin, « *Réduction de la dimension des données* », Université de Sherbrooke, Sherbrooke, 2020, pp.21-24.
- [7]. Bounneche Meriem Dorsaf, «*Réduction de données pour le traitement d'images*», Mémoire de magistère option contrôle, Université Mentouri Constantine, 2009, pp.87.
- [8]. S. K. Hese, M. R. Banwaskar, «Performance Evaluation of PCA and LDA for Face Recognition», International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Volume 2, Issue 2, February 2013, pp. 149-154.
- [9]. Önsen TOYGAR, Adnan ACAN, «Face Recognition Using PCA, LDA and ICA Approaches On Colored Images », Istanbul University – Journal of Electrical & Electronics Engineering, Volume 3 , Numéro 1, 2003, pp. 735-743.
- [10]. Nasserallah allam, *Application de l'analyse discriminante linéaire (LDA) dans la reconnaissance du visage*, mémoire de Master en Informatique, Université Kasdi-Merbeah Ouargla, 23Juin 2018, pp. 62.
- [11]. Arunasakthi. K,KamatchiPriya. L, «A Review on Linear and Non-linear Dimensionality Reduction Techniques», Machine Learning and Applications: An International Journal (MLAIJ) Vol.1, No.1, September 2014, pp.12.
- [12]. Mr. GHALI Ahmed, *Amélioration de la reconnaissance par le visage*, mémoire de Magister en Simulation et Intelligence Artificielle, Université des sciences et de la technologie d'Oran Mohamed Boudiaf, Oran, 2015, pp. 107.
- [13]. BOUDJELLAL Sofiane, « Détection et identification de personne par méthode biométrique », Mémoire de Magister en Electronique, Université Mouloud MAMMARI de TIZI-OUZOU (UMMTO), pp. 95.
- [14]. Wei-Lun Chao, *Dimensionality Reduction*, Graduate Institute of Communication Engineering, National Taiwan University, 2011, pp. 57.
- [15]. *Capítulo 13 Discriminant Analysis*,
- [17]. Peter N. Belhumeur, Joao P. Hespanha et David J. Kriegman, «Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection », 1996 European conference on computer vision, 1996, pp.15.

- [18]. Mr. Rahul M. Ohol<sup>1</sup>, Mrs. Shilpa R. Ohol<sup>2</sup>, *Linear Discriminant Analysis for Human Face Recognition*, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 08, 2017, pp.1005.
- [19]. Ricardo Gutierrez-Osuna, *LECTURE 10: Linear Discriminant Analysis*, Introduction to Pattern Analysis Texas A&M University, pp.15
- [20]. Gaber, T, Tharwat, A, Ibrahim, A et Hassanien, AE, «Linear discriminant analysis: a detailed tutorial», 2017, pp. 23.
- [21]. Jie Yang, Hua Yu, William Kunz, «An Efficient LDA Algorithm for Face Recognition», School of Computer Science Interactive Systems Laboratories, Pittsburgh, PA 15213, pp.6.
- [22]. Kuldeep Singh Sodhi et Madan Lal, *Face recognition PCA, LDA, and Various Distance Classifiers*, Journal of Global Research in Computer Science, Volume 4, No. 3, Mars 2013, pp. 30-35.
- [23]. Juwei Lu, *Discriminant Learning for Face Recognition*, thèse de doctorat de philosophie, université de Toronto, Toronto, 2004, pp. 175.
- [24]. Chiheb Amira et Bouhalit nasereddine, «<< Reconnaissance de visages par LDA >>», Mémoire de licence en informatique, sous la direction de [Mlle Zedadra], Guelma, Université de 8 mai 45 de Guelma Algérie, 2003, 37 p.
- [25]. Su-Yun Huang, Chii-Ruey Hwang, Miao-Hsiang Lin, Kernel Fisher's Discriminant Analysis in Gaussian Reproducing Kernel Hilbert Space, Feb. 1, 2005, pp. 35.
- [26]. Masashi Sugiyama, *Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction*, Tokyo Institute of Technology, 2-12-1-W8-74, O-okayama, Meguro-ku, pp. 8.
- [27]. Line Clemmensen, Trevor Hastie, Daniela Witten, Bjarne Ersbøll, *Sparse Discriminant Analysis*, 16 avril 2011, pp. 21.
- [28]. Yi Wu, David Wipf, Jeong-Min Yun, *Understanding and Evaluating Sparse Linear Discriminant Analysis*, pp. 1070-1078.
- [29]. M me CHAIBI Sonia, *Accélération de l'Apprentissage des SVMs pour le traitement de bases de données de grandes tailles : Application à la reconnaissance des mots arabes isolés*, thèse de doctorat en sciences, Université Badji Mokhtar-Annaba, 2015, pp. 141.
- [30]: Amritpal Kaur, Sarabjit Singh et Taqdir, «<< Face Recognition Using PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) Techniques >>», International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 3, Mars 2015, pp. 308-310.
- [31]. Kuldeep Singh Sodhi et Madan Lal, *FACE RECOGNITION USING PCA, LDA AND VARIOUS DISTANCE CLASSIFIERS*, Journal of Global Research in Computer Science, Volume 4, No. 3, Mars 2013, pp. 30-35.
- [32]. Mr. Rahul M. Ohol<sup>1</sup>, Mrs. Shilpa R. Ohol<sup>2</sup>, *Linear Discriminant Analysis for Human Face Recognition*, International Research Journal of Engineering and Technology (IRJET), Volume: 04, issue: 08, Août 2017, pp. 1005-1007.
- [33]. MERAMRIA Nabila, «<< Reconnaissance de visages par Analyse Discriminante Linéaire(LDA)>>», Mémoire de master en : Télécommunications Avancées, sous la direction de [Zermi .N], Annaba, UNIVERSITE BADJI MOKHTAR ANNABA, 2016, 52 p.
- [34]. Thales Group «<< la biométrie au service de l'identification et l'authentification>>», 14 avril 2021,

- [35]. Nyemo Kodjovi Koumadi, << Authentification Automatique du propriétaire d'un téléphone mobile >>, Mémoire de en vue de l'obtention du grade de maitre de maitre en informatique, Québec, UQAC Université de Québec à Chicoutimi, Juillet 2018, 83 p.
- [36]. Shireen Elhabian et Aly Farag, *2D Face Recognition Using PCA, ICA and LDA*, CIVP Lab, Université de Louisville, Octobre 2009, 30 p.
- [37]. Nicolas MORIZET, << Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris >>, thèse de doctorat en spécialité Signal et Images, sous la direction de [M. Amara Amara], Paris, École Doctorale d'Informatique, Télécommunications et Électronique de Paris, 18 Mars 2009, 211p.
- [38]. Commissariat à la protection de la vie privée du Canada, *Reconnaissance faciale automatisée dans les secteurs public et privé*, Québec, Mars 2013,
- [39]. Mustamin Anggo et La Arapu, *Face Recognition Using Fisherface Method*, 2018 J. Phys.: Conf. Ser. 1028 012119, 2018, pp. 1-9.
- [40]. Id3 Technologies, *Reconnaissance faciale : la science de l'intelligence artificielle pour l'identification biométrique des personnes par le visage*, Fontanil-Camillon France, 2021,
- [41]. Gasmi Lynda, *Deep learning for face recognition*, memoire de master en informatique, université Mohamed Boudiaf-M'sila, 2020, pp. 66.
- [42]. Alaa, Hilal, système d'identification à partir de l'image d'iris et détermination de la localisation des informations, thèse de doctorat en optimisation et sureté des systèmes, université libanaise de Beyrouth, 2013 pp. 164.
- [43]. <https://www.programmingsought.com/article/57516306132/>
- [44]. Chiheb Amira et Bouhalit nasereeddine, << Reconnaissance de visages par LDA >>, Mémoire de licence en informatique, sous la direction de [Mlle Zedadra], Guelma, Université de 8 mai 45 de Guelma Algérie, 2003, 37 p.
- [45]. John P. Cunningham, Zoubin Ghahramani, *Linear Dimensionality Reduction: Survey, Insights, and Generalizations*, journal of Machine Learning Research 16 (2015) 2859-2900 Submitted 5/14; Revised 3/15; Published 12/15, Cambridge, pp. 2860-2900.
- [46]. Kamran Etemad and Rama Chellappa, Discriminant analysis for recognition of human face images, Department of Electrical Engineering and Center for Automation Research, University of Maryland, Vol. 14, No. 8/ August 1997, pp. 1724-1733.
- [47]. Juwei Lu, Konstantinos N. Plataniotis, Konstantinos N. Plataniotis, Face Recognition Using Kernel Direct Discriminant Analysis Algorithms, IEEE Transactions on Neural Networks, vol. 14, No. 1, Janvier 2003, pp. 117-126.
- [48]. Laboratoire AT&T Cambridge, URL: <https://cam-orl.co.uk/facedatabase.html>
- [49]. <https://www.face-rec.org/databases/>
- [50]. [https://git-disl.github.io/GTDLBench/datasets/att\\_face\\_dataset/](https://git-disl.github.io/GTDLBench/datasets/att_face_dataset/)

[51]. Sohini Roychowdhury<sup>1</sup> and Michelle Emmons, *A Survey Of The Trends in Facial and Expression Recognition Databases and Methods*, pp. 16.

[52]. <https://fr.mathworks.com/matlabcentral/fileexchange/35106-the-phd-face-recognition-toolbox>