

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université Mohamed Khider Biskra  
**Faculté des Sciences Exactes, des sciences de la Nature et de la Vie**  
**Département d'Informatique**

N° d'ordre :.....

Série :.....



## **Mémoire**

Présenté en vue de l'obtention du diplôme de Magister en Informatique  
Option: **Intelligence artificielle et systèmes distribués**

Titre :

# **Reconnaissance d'objets dans les images à base de SVM (application à la détection de visages)**

Par :

**ZEGHICHI Amel**

Soutenu le :

**Devant le jury :**

<b>Pr. N. DJEDI</b>	<b>PR</b>	<b>Université de Biskra</b>	<b>Président</b>
<b>Dr. M.C.BABAHENINI</b>	<b>MCA</b>	<b>Université de Biskra</b>	<b>Rapporteur</b>
<b>Dr. F. CHERIF</b>	<b>MCA</b>	<b>Université de Biskra</b>	<b>Examineur</b>
<b>Dr. A. MOUSSAOUI</b>	<b>MCA</b>	<b>Université de Setif</b>	<b>Examineur</b>

*à tous ceux que j'aime,  
à tous ceux qui m'aiment,  
et à tous ceux qui m'ont aidé, encouragé,  
et soutenu de loin ou de près dans les  
moments difficiles*

# Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à monsieur Babahenini Mouhamed Chaouki de m'avoir encadrée et orientée, pour ses conseils, sa confiance et son assistance tout au long de ces années de mémoire. Je le remercie également pour sa disponibilité, sa compréhension et sa relecture minutieuse du mémoire.

Ma plus grande reconnaissance à monsieur Djedi Nourinne pour avoir accepté de présider mon jury de mémoire. Je remercie aussi monsieur Foudil Cherif d'avoir examiné mon travail avec un grand intérêt. Mes sincères remerciements à monsieur Moussaoui Abdelouahab et mon profond respect pour son éventuel déplacement afin d'être présent parmi les membres du jury.

Je remercie énormément ma famille et surtout mère qui m'a encouragé et soutenu toute ma vie estudiantine et d'avoir été patiente pendant ces dernières années.

Enfin toute ma sympathie à mes collègues et à mes chères amis

---

# Résumé

---

Durant la dernière décennie la haute sécurité est devenue un problème majeur pour les installations gouvernementales (militaire, nucléaire, aéroport, ..). L'identification de personne par la reconnaissance de visages donne des solutions à ce problème. Nous présentons dans ce mémoire un système d'identification de visage basé sur l'utilisation des descripteurs SIFT et la méthode SVM. Il s'agit tout d'abord de détecter sur l'image des zones circulaires « intéressantes », centrées autour d'un *point-clé* et de rayon déterminé appelé *facteur d'échelle*, celles-ci sont caractérisées par leur unité visuelle et correspondent en général à des éléments distincts sur l'image. Sur chacune d'elles, on détermine une orientation intrinsèque qui sert de base à la construction d'un histogramme des orientations locales des contours, habilement pondéré, seuillé et normalisé pour plus de stabilité. Cet histogramme, sous la forme d'un vecteur à 128 dimensions (ou valeurs) constitue le descripteur SIFT du point-clé, et l'ensemble des descripteurs d'une image établissent ainsi une véritable signature numérique de son contenu. Les descripteurs SIFT présentent l'avantage d'être invariants à l'orientation et à la résolution de l'image, et peu sensibles à son exposition. Une fois le calcul des descripteurs effectué, ils seront utilisés comme entrée au module SVM qui va générer un modèle de décision pour chaque classe en utilisant la méthode "un contre tous". Ce modèle sera utilisé pour décider à propos des nouvelles images.

**Mots-clés:** Reconnaissance de visages, Machine à vecteur de support (SVM), transformation de caractéristiques visuelles invariante à l'échelle (SIFT), facteur d'échelle.

---

# Abstract

---

In the last decade the high security has become a major problem for government facilities (military, nuclear, airport ...). Identification of person by face recognition gives solutions to this problem. We present in this work a face recognition system based on SIFT descriptors and SVM method. First, we detected on the image circle areas "interesting", centered on a key point and fixed radius called scale factor. These are characterized by their visual unity and generally correspond to distinct elements in the image. On each one, we determine an intrinsic orientation which is the basis for constructing a histogram of local orientations of the contours, deftly weighted thresholded and normalized for stability. This histogram as a vector of 128 dimensions (or values) of the SIFT descriptor is the key point, and all the descriptors of an image establish a true digital signature of its content. The SIFT descriptors have the advantage of being invariant to the orientation, image resolution, and insensitive to exposure. Once the calculation of descriptors is made, they are used as input to the SVM module that will generate a decision model for each class using the method "one against all". This model will be used to decide about the new images.

**Keywords:** Face recognition, support vector machine (SVM), Scale Invariant Feature Transform (SIFT), scale factor.

# ملخص

في العشرية الأخيرة، أصبح الأمن يمثل مشكلة رئيسية للمرافق الحكومية (النوعية، العسكرية، الموانئ، و المطارات...)، لهذا السبب فإن تحديد هوية الشخص عن طريق التعرف على الوجه يعطي عددا من الحلول لهذه المشكلة. نقدم في هذه المذكرة نظام التعرف على الوجه باستخدام الوصف SIFT وطريقة SVM. في البداية نقوم باكتشاف مجالات دائرية "ذات الأهمية" في الصورة، و التي تتمحور حول نقطة رئيسية ونصف قطرها ثابت يسمى عامل الحجم. تتميز هذه المجالات بالوحدة المرئية و تتوافق عموما مع عناصر مختلفة في الصورة. على كل واحد منها، نقوم بتحديد التوجه العام الذي هو أساس لبناء رسم بياني من التوجهات المحلية للحدود. هذا الرسم البياني الممثل بشعاع من 128 بعد (أو قيمة) يمثل الوصف SIFT للنقطة الرئيسية، وجميع الواصفات للصورة تؤسس التوقيع الرقمي الصحيح لمضمونها. الواصفات SIFT لديها ميزة كونها ثابتة بالنسبة للإتجاه و دقة الصورة و حجمها. بعد احتساب الوصف SIFT يتم استخدامه كمدخل إلى وحدة SVM التي من شأنها أن تولد نموذج قرار باستخدام أسلوب "واحد ضد الجميع". يتم استخدام هذا النموذج لاتخاذ قرار حول الصور الجديدة.

**الكلمات المفتاحية:** التعرف على الوجوه، جهاز الدعم الموجه (SVM)، تحول الخصائص البصرية الثابتة بالنسبة الحجم (SIFT)، عامل الحجم.

# Table des matières

<b>Introduction général</b> .....	1
<b>Chapitre I : les systèmes biométriques</b>	
1. Introduction.....	5
2. Définition.....	5
3. Les propriétés des caractéristiques biométriques.....	6
4. Architecture d'un système biométrique.....	7
5. Présentation des différentes technologies biométries.....	8
5.1. Empreintes digitale.....	8
5.2. Géométrie de la main.....	9
5.3. Visage.....	10
5.4. La rétine.....	11
5.5. L'iris.....	12
5.6. Voix.....	12
5.7. Signature dynamique.....	13
5.8. Dynamique de la frappe au clavier.....	14
6. Evaluation d'un système biométrique.....	15
7. La fiabilité des systèmes biométriques.....	17
8. Les applications de la biométrie.....	18
8.1. Contrôle d'accès .....	18
8.1.1. Contrôle d'accès physique.....	18
8.1.2. Contrôle d'accès virtuel.....	19
8.2. Authentification des transactions .....	19
8.3. Répression .....	20
8.4. Personnalisation .....	20
9. La place de la reconnaissance faciale parmi les autres techniques biométriques .....	20
10. Conclusion .....	21

## **Chapitre II : La reconnaissance des visages humains.**

1.	Introduction.....	23
2.	Motivation : (pourquoi la reconnaissance de visages ?).....	24
3.	Qu'est ce qu'un système de reconnaissance de visage ?.....	25
3.1.	Le monde physique .....	26
3.2.	Acquisition de l'image .....	26
3.3.	Prétraitement .....	26
3.4.	Extraction de caractéristiques.....	26
3.5.	Apprentissage.....	27
3.6.	Décision.....	27
4.	Difficultés inhérentes à la reconnaissance automatique de visage .....	27
4.1.	Changement d'illumination .....	27
4.2.	les variations de pose .....	28
4.3.	Occlusions partielles .....	29
4.4.	Expressions faciales .....	29
4.5.	Présence ou absence des composants structurels .....	30
4.6.	L'impact de la taille de la base .....	30
5.	Les méthodes de reconnaissance de visage .....	30
5.1.	Les méthodes locales.....	30
5.2.	Les méthodes globales.....	31
5.3.	Les méthodes hybrides.....	31
6.	Les algorithmes classiques de reconnaissance de visages .....	32
6.1.	Analyse en composantes principales .....	32
6.2.	Les réseaux de neurones .....	34
6.3.	Les machines à vecteur de supports .....	35
6.4.	Template matching .....	36
6.5.	le modèle de Markov caché .....	36
6.6.	l'utilisation des points caractéristiques .....	37
6.7.	Appariement élastique des graphes .....	38
7.	Comparaison des algorithmes .....	39
8.	Conclusion.....	40

## Chapitre III : Les machines à vecteurs de support.

1.	Introduction.....	41
2.	Aperçu sur les SVM.....	41
2.1.	Théorie de base des SVM .....	45
2.2.	Classification des données multi classe .....	50
3.	Technique de reconnaissance de visages à base de SVM.....	56
3.1	NEAR SET et SVM pour la reconnaissance de visage.....	45
3.2	Caractéristiques de Gabor et SVM pour la reconnaissance de visage.....	46
3.3	PCA, LDA et SVM pour la reconnaissance de visage.....	46
3.4	Transformée d'ondelettes et SVM pour la reconnaissance de visage.....	47
3.5	KPCA et SVM pour la reconnaissance de visage.....	47
3.6	Bayesian et SVM pour la reconnaissance de visage.....	48
3.7	Ondelette, ACP et SVM pour la reconnaissance de visage.....	48
3.8	Vissage propres et SVM pour la reconnaissance de visage.....	49
3.9	GMM et SVM pour la reconnaissance de visage.....	49
7.	Conclusion.....	50

## Chapitre IV : Approche hybride pour la reconnaissance de visages à base de SIFT et SVM

1.	Introduction et motivation.....	51
2.	L'architecture globale du Système .....	52
2.1.	Calcul des points d'intérêt et de leurs descripteurs SIFT.....	53
2.1.1.	Détection d'extremums d'espace-échelle (scale-space).....	53
2.1.2.	Localisation des points d'intérêt .....	56
2.1.2.1.	Teste d'élimination des points-clés.....	57
2.1.2.1.1.	Les faibles réponses.....	57
2.1.2.1.2.	Les bords.....	58
2.1.3.	Choix de l'orientation des descripteurs.....	59
2.1.4	Calcul des descripteurs.....	60
2.2	Apprentissage (SVM).....	62
2.3	Mise en correspondance des points d'intérêts.....	63
2.4	Classification par SVM.....	65
3.	Conclusion.....	65

## Chapitre V : Mise en œuvre, bilans et résultats.

1.	Introduction.....	67
2.	Réalisation du système.....	67
2.1.	SIFT.....	68
2.2.	SVM.....	71
3.	Tests et Expérimentations.....	73
3.1.	Base de données .....	73
3.2.	SIFT.....	75
3.3.	SVM.....	76
3.3.1.	Noyau linéaire.....	76
3.3.2.	Noyau polynomial.....	77
3.3.3.	Noyau RBF.....	79
4.	Bilan globale et Conclusion.....	82
	<b>Conclusion et perspectives.....</b>	<b>83</b>

## Annexe A: Les bases de données utilisées

1.	La base FERET .....	85
2.	La base de Yale.....	86
3.	La base ORL.....	86
4.	La base PF01.....	87
5.	La base UMIST.....	88
6.	La base AR.....	89
7.	La base PIE.....	90
8.	La base VISIO.....	91

## Annexe B: Dualité Lagrangienne et passage à la forme duale

1.	Introduction.....	94
2.	Dualité Lagrangienne .....	94
3.	Transformation du problème d'apprentissage en forme duale.....	100
3.1.	Cas linéairement séparables.....	100
3.2.	Cas de données non linéairement séparables.....	103
3.2.1.	Cas des fonctions noyaux.....	103
3.2.2.	Cas de l'introduction des variables de pénalités.....	103

	<b>Bibliographie .....</b>	<b>106</b>
--	----------------------------	------------

# Liste des figures

FIG 1.	Schématisation du processus de reconnaissance de visage.....	3
FIG 2.	Principaux modules d'un système biométrique ainsi que les différentes modes....	7
FIG 3.	le processus de reconnaissance par empreinte digitale.....	8
FIG 4.	dispositif de reconnaissance par géométrie de la main.....	9
FIG 5.	Photo de rétine.....	11
FIG 6.	Photo d'iris.....	12
FIG 7.	Spectre d'un signal voix.....	13
FIG 8.	Signature.....	14
FIG 9.	les différents critères pour chaque type de technique biométrique.....	16
FIG 10.	La recherche d'un compromis entre les 2 taux FFR et FAR résulte le point X, la zone delta représente la marge d'erreur autorisée par le système.....	18
FIG 11.	Schéma générale d'un système de reconnaissance de visages.....	25
FIG 12.	Extrait de [28] Effets de variations dans les directions d'illumination sur l'apparence d'un visage.....	28
FIG 13.	Exemple de variation de pose.....	29
FIG 14.	Exemple de variation d'expression.....	29
FIG 15.	Modèle de visage composé de 16 régions (les rectangles) associées à 23 relations (flèches) [48].....	36
FIG 16.	Modèle géométrique de visage.....	38
FIG 17.	Structure des approches EGM et EBGm.....	39
FIG 18.	Classification entre deux classes.....	42
FIG 19.	Illustration de l'effet de changement d'espace par une fonction .....	44
FIG 20.	Architecture globale du système .....	52
FIG 21.	Pyramide de gradients: 3 octaves de 5 gradients.....	54
FIG 22.	Construction de la fonction d'analyse multi-échelle de l'image .....	55
FIG 23.	Construction de la Pyramide de différences de gaussiens (DoG) à partir de la pyramide de gradients .....	55
FIG 24.	Comparaison aux 26 plus proches voisins.....	56
FIG 25.	Exemple de détection d'extremums dans l'espace des échelles.....	56
FIG 26.	Construction de l'histogramme des orientations.....	60

FIG 27.	Construction d'un descripteur SIFT.....	60
FIG 28.	Apprentissage(SVM) .....	63
FIG 29.	Mise en correspondance des points d'intérêts.....	64
FIG 30.	Classification par SVM.....	65
FIG 31.	Des Exemples de la base de données ORL.....	74
FIG 32.	3 exemples d'images.....	75
FIG 33.	Les images avec les descripteurs SIFT.....	75
FIG 34.	Courbe représentant le taux de reconnaissance du noyau polynomial en utilisant A1 et T1.....	77
FIG 36.	Courbe représentant le taux de reconnaissance du noyau polynomial en utilisant A3 et T3.....	78
FIG 37.	Courbe représentant le taux de reconnaissance du noyau RBF en utilisant A1 et T1.....	79
FIG 38.	Courbe représentant le taux de reconnaissance du noyau RBF en utilisant A2 et T2.....	80
FIG 39.	Courbe représentant le taux de reconnaissance du noyau RBF en utilisant A3 et T3.....	80
FIG 40.	Extrait de la base FERET.....	85
FIG 41.	Extrait de la base Yale.....	86
FIG 42.	Extrait de la base ORL.....	87
FIG 43.	Extrait de la base PF01.....	88
FIG 44.	Extrait de la base UMIST.....	89
FIG 45.	Extrait de la base AR.....	90
FIG 46.	Extrait de la base PIE: variation de pose.....	91
FIG 47.	Extrait de la base PIE: à gauche des vues Illumination I et à droite des vues Illumination II.....	91
FIG 48.	Variation des images pris pour une personne dans la base VISIO.....	93

# Liste des tableaux

Tableau 1	Répartition des articles sur les techniques biométriques soumis et acceptés dans la revue IEEE PAMI.....	21
Tableau 2.	Comparaison entre les algorithmes de reconnaissance de visage.....	40
Tableau 3.	Comparaison entre l'approche SIFT et Eigenfaces .....	75
Tableau 4.	les résultats obtenus en utilisant le noyau linéaire.....	76
Tableau 5.	les meilleurs résultats obtenus pour chaque noyau.....	81
Tableau 6.	les meilleurs résultats obtenus pour chaque Méthode .....	81
Tableau 7.	Principales Caractéristiques des bases des visages [75].....	84
Tableau 8.	Les variations dans la base de données VISIO.....	92

# Introduction Générale

## *Contexte*

Savoir déterminer de manière à la fois efficace et exacte l'identité d'un individu est devenu un problème critique dans notre société. En effet, bien que nous ne rendions pas toujours compte, notre identité est vérifiée quotidiennement par de multiples organisations: lorsque nous utilisons notre carte bancaire, lorsque nous accédons à notre lieu de travail, lorsque nous nous connectons à un réseau informatique, etc...

Il existe traditionnellement deux manières d'identifier un individu. La première méthode est basée sur une connaissance (knowledge-based), cette connaissance correspond par exemple au mot de passe utilisé au démarrage d'une session Unix ou bien au code qui permet d'activer un téléphone portable. La seconde méthode est basée sur une possession (token-based), il peut s'agir d'une pièce d'identité, une clef, un badge, etc... Ces deux modes d'identification peuvent être utilisés de manière complémentaire afin d'obtenir une sécurité accrue; cependant, elles ont leurs faiblesses respectives. Dans le premier cas, le mot de passe peut être oublié par son utilisateur ou bien deviné par une autre personne et dans le second cas, le badge (ou la pièce d'identité ou la clef) peut être perdu ou volé.

La biométrie est une alternative aux deux précédents modes d'identification. Elle consiste à identifier une personne à partir de ses caractéristiques physiques ou comportementales. Le visage, les empreintes digitales, l'iris, etc.. sont des exemples de caractéristique physiques. La voix, l'écriture, le rythme de frappe sur un clavier, etc... sont des caractéristiques comportementales. Ces caractéristiques sont attachées à chaque individu et ne souffrent donc pas des faiblesses des méthodes basées sur une connaissance ou une possession. En général, elles sont très difficiles à deviner ou à voler ainsi qu'à dupliquer [1].

L'une des branches de la vision par ordinateur qui remporte un grand succès et qui est en perpétuel développement est la reconnaissance de visage. Le visage est

certainement la caractéristique biométrique que les humains utilisent le plus naturellement pour s'identifier entre eux. Le but d'un système de reconnaissance de visage est de simuler le système de reconnaissance humain par la machine pour automatiser certaines applications telles que : la télésurveillance, le contrôle d'accès à des sites, l'accès à des bâtiments sécurisés, etc....

La difficulté de la reconnaissance de visage par ordinateur varie énormément suivant les conditions d'acquisition. Dans un environnement contrôlé, des paramètres tels que l'arrière plan, la direction et l'intensité des sources lumineuses, l'angle de la prise de vue, la distance de la caméra au sujet sont maîtrisés par le système. Dans un environnement non contrôlé, une série de prétraitements sont souvent nécessaires avant de faire la reconnaissance proprement dite.

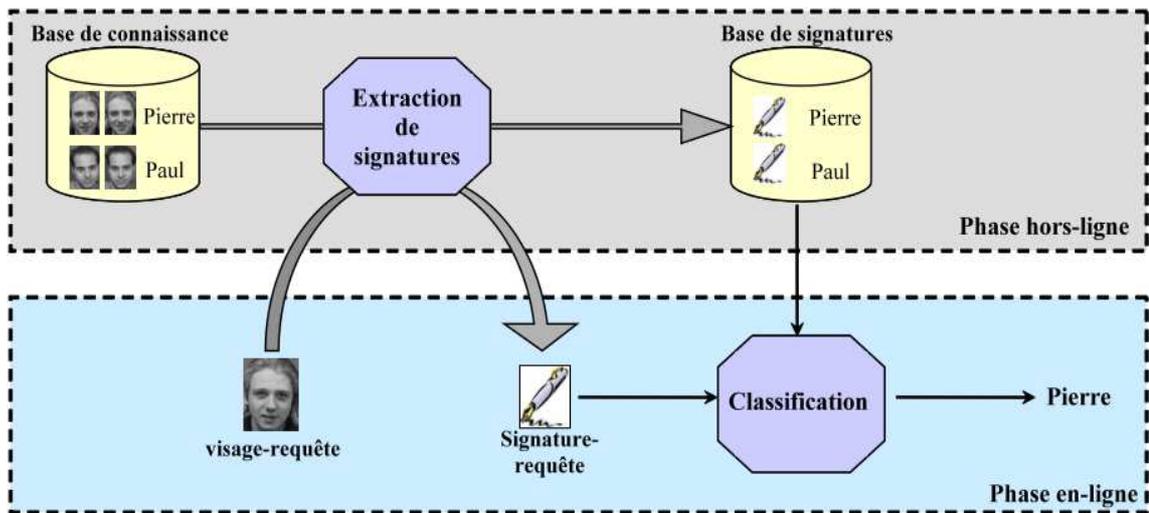
Actuellement il existe de nombreuses méthodes qui permettent de reconnaître et d'identifier une personne dans une image. On peut diviser ces méthodes en deux catégories: les méthodes géométriques, et les méthodes globales. La performance de ces méthodes dépend de la précision (nombre de paramètres) avec laquelle les informations utiles du visage sont extraites (comme certaines partie du visage les yeux, le nez, la bouche, ...) [2] [3] [4] [5].

### ***Problématique***

Comme toute tâche de reconnaissance d'objets, un système automatique de reconnaissance de visage est composé de deux étapes importantes (figure I):

- ✓ ***Extraction d'élément caractéristique (signature):*** pour le choix de signature, on privilégie un certain nombre de propriétés. Celles-ci doivent être invariantes et discriminantes. Le terme invariant s'applique à des caractéristiques qui ne sont pas ou peu affectées par des changements acceptables d'apparence. En d'autres termes, les signatures extraites de deux vues d'une même personne doivent être le plus proche possible, même si les conditions de prise de vue sont différentes. Le terme discriminant indique que ces caractéristiques doivent, de plus, prendre des valeurs significativement différentes pour les vues de deux personnes distinctes.
- ✓ ***Classification des signatures:*** l'ensemble des images des visages connus sont stockés dans une base de connaissance. Cette base de connaissance peut contenir plusieurs images d'une même personne, sous des conditions de prise de vue différentes. Chaque

image est étiquetée par son identité associée. A chaque personne de la base de connaissance, on associe une signature qui lui est caractérisée. La reconnaissance d'un visage requête se fait en deux étapes. Dans une première phase, on extrait sa signature à l'aide de la même technique que celle appliquée à la base d'apprentissage; puis, la signature requête ainsi obtenue est mise en correspondance avec la signature de la personne la plus proche dans la base de connaissance. On en déduit l'identité du visage requête.



**FIG 1:** Schématisation du processus de reconnaissance de visage

### *Contributions*

Dans ce mémoire, nous proposons un modèle pour la reconnaissance automatique de visage, combinant entre deux approches: le scale invariant feature transform pour l'extraction des signatures caractéristiques et les machines à vecteurs de supports pour la classification de celles-ci. Chaque image représentant un visage est décrite par un vecteur caractéristique composé de 70 descripteurs. Ces descripteurs sont associés à l'orientation et l'échelle à laquelle ils ont été détectés. Les points sur lesquels ces descripteurs sont calculés sont choisis en filtrant l'image par un filtre Gaussien à différentes échelles puis en extrayant les extrema en espace et en échelle. Les vecteurs caractéristiques calculés sont ensuite utilisés comme entrée au module SVM qui va générer un modèle de décision pour chaque classe en utilisant la méthode "un contre tous".

## *Organisation du mémoire*

Ce mémoire se compose de cinq grands chapitres et il est organisé de la manière suivante:

La première partie "*Etat de l'art*" composée de quatre chapitres:

✚ Le premier chapitre est consacré à la présentation générale de la biométrie. Il décrit tout d'abord le principe de fonctionnement des systèmes biométriques puis définit les outils utilisés pour évaluer leurs performances. Ensuite, la place de la reconnaissance faciale parmi les autres techniques biométriques est analysée.

✚ Dans le deuxième chapitre nous présentons le processus de reconnaissance de visage et étudions les principales techniques proposées. Nous présentons les performances de ces techniques, et précisons leurs avantages et inconvénients.

✚ Dans le troisième chapitre nous présentons une description de la méthode de classification SVM (support vecteur machine), puis nous introduisons une relecture de quelques travaux récents relevant de ce domaine.

La deuxième partie "*Contribution*" composée de deux chapitres:

✚ Dans le quatrième chapitre nous décrivons la méthode proposée basée sur les descripteurs SIFT et les machines à vecteurs de supports.

✚ Dans le cinquième chapitre sont présentés l'implémentation du système et les résultats expérimentaux.

Une conclusion générale terminera ce mémoire et introduira quelques perspectives sur les travaux futures.

# Chapitre I

## Les systèmes biométriques

### 1. Introduction

A ce jour, le moyen le plus répandu d'authentifier une personne de manière électronique repose sur l'utilisation d'un code personnel, composé de chiffres et/ou de lettres. Ce mode de vérification d'identité est relativement peu sûr, puisqu'il suffit de connaître le code de quelqu'un d'autre pour usurper son identité. De plus, la multiplication des transactions et des communications électroniques fait que l'utilisateur doit retenir un nombre croissant de codes: un pour retirer de l'argent depuis un distributeur automatique, un pour accéder à son ordinateur, plusieurs pour se connecter à des sites internet, etc. Si bien que le besoin de systèmes d'authentifications conviviales et sûres se fait de plus en plus pressant. La biométrie consiste à authentifier une personne à l'aide de ce qu'elle est, et non pas de ce qu'elle sait ou détient.

### 2. Définition

Un système biométrique est un système de reconnaissance de personne sur la base de caractères physiologiques ou de traits comportementaux automatiquement reconnaissables et vérifiables [6]. L'identification biométrique est basée sur l'analyse de données liées à l'individu et peut être classée en trois grandes catégories :

- L'analyse morphologique. (empreinte digitale, forme de la main, les traits du visage, réseau veineux de la rétine, iris de l'œil, voix, etc.)
- Les traces biologiques. (odeur, salive, urine, sang, ADN, etc.)

- L'analyse comportementale. (dynamique du tracé de signature, frappe sur un clavier d'ordinateur.

Selon le contexte de l'application, un système biométrique peut fonctionner en mode d'enrôlement, en mode de vérification ou bien en mode d'identification [7] :

- Le mode d'enrôlement est une phase d'apprentissage qui a pour but de recueillir des informations biométriques sur les personnes à identifier. Plusieurs campagnes d'acquisitions de données peuvent être réalisées afin d'assurer une certaine robustesse au système de reconnaissance aux variations temporelles des données. Pendant cette phase, les caractéristiques biométriques des individus sont saisies par un capteur biométrique, puis représentées sous forme numérique (signatures), et en fin stockées dans la base de données. Le traitement lié à l'enrôlement n'a pas de contrainte de temps, puisqu'il s'effectue « hors-ligne ».
- Le mode de vérification ou authentification est une comparaison "un à un", dans lequel le système valide l'identité d'une personne en comparant les données biométriques saisie avec le modèle biométrique de cette personne stockée dans la base de données du système. Dans un tel mode, le système doit alors répondre à la question suivante: «*Suis-je réellement la personne que je suis en train de proclamer?*». Actuellement la vérification est réalisée via un numéro d'identification personnel, un nom d'utilisateur, ou bien une carte à puce.
- Le mode d'*identification* est une comparaison "un à N", dans lequel le système reconnaît un individu en l'appariant avec un des modèles de la base de données. La personne peut ne pas être dans la base de données. Ce mode consiste à associer une identité à une personne. En d'autres termes, il répond à des questions du type: « *Qui suis-je ?* ».

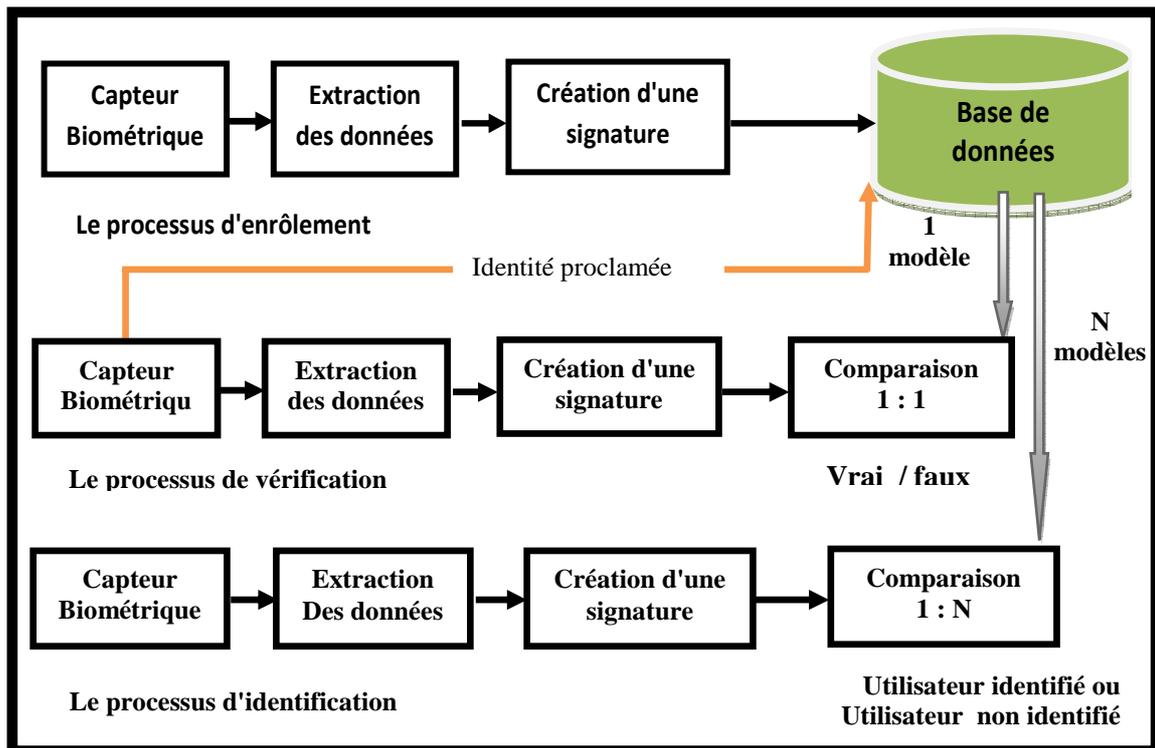
### 3. Les propriétés des caractéristiques biométriques

Une caractéristique biométrique doit être [8]:

- Universelle, c'est-à-dire que toutes les personnes de la population à identifier doivent la posséder.
- A la fois facilement et quantitativement mesurable.

- Unique, c'est-à-dire que deux personnes ne peuvent pas posséder exactement la même caractéristique.
- Permanente, ce que signifie qu'elle ne doit pas varier au cours du temps.
- Performante, c'est-à-dire que l'identification doit être précise et rapide.
- Bien acceptée par les utilisateurs du système.
- Impossible à dupliquer par un imposteur.

#### 4. Architecture d'un système biométrique



**FIG 2:** Principaux modules d'un système biométrique ainsi que les différents modes [7].

Les différents modules qui composent un système biométrique sont représentés sur la figure 2 ; leur fonctionnement peut être résumé comme suit [7]:

- **Module capteur biométrique:** correspond à la lecture de certaines caractéristiques physiologiques, comportementales ou biologiques d'une personne, au moyen d'un terminal de capture biométrique (ou capteur biométrique);
- **Module extraction des données:** extrait les informations pertinentes à partir des données biométriques brutes, par exemple des images de visage ou des régions caractéristiques de visage ;

- **Module création d'une signature**: crée un modèle numérique afin de représenter la donnée biométrique acquise. Ce modèle, appelé aussi signature, sera conservé sur un support portable (puce ou autre) ou dans une base de données ;
- **Module comparaison** : compare les caractéristiques biométriques d'une personne soumise à contrôle (volontairement ou à son insu) avec les « signatures » mémorisées. Ce module fonctionne soit en mode vérification (pour une identité proclamée) ou bien en mode identification (pour une identité recherchée).
- **Module base de données** : stocke les modèles biométriques des utilisateurs enrôlés.

## 5. Présentation des différentes technologies biométries

### 5.1. Empreintes digitales

La reconnaissance des empreintes digitales est la technique biométrique la plus ancienne et c'est l'une des plus matures. Elle se base sur le fait que chaque personne a des empreintes uniques. Après la capture de l'image de l'empreinte, on fait un rehaussement de l'image. Ensuite on identifie et on extrait les minuties, qui vont être comparées avec l'ensemble des minuties sauvegardées des autres utilisateurs. C'est l'une des technologies biométriques les plus étudiées et les plus utilisées, surtout dans le contrôle d'accès [6].



FIG 3 : le processus de reconnaissance par empreinte digitale

Nous avons essayé de résumer dans le tableau ci-dessous les principaux avantages et les limites essentielles de cette technique :

Avantages	Inconvénients
<ul style="list-style-type: none"><li>• Son ancienneté et sa mise à l'épreuve</li><li>• Sa résistance aux changements de température, jusqu'à un certain point.</li><li>• Sa capacité à fournir des résolutions de plus de 500 dpi.</li></ul>	<ul style="list-style-type: none"><li>• Possibilité d'endommagement ou d'élimination des empreintes.</li><li>• Le dispositif CCD peut s'user avec le temps et devenir moins fiable.</li><li>• Problèmes de contrastes (un doigt propre et sec devient trop clair tandis qu'un doigt humide et recouvert d'un film gras devient très foncé), problème résolu grâce au film liquide mais système mal accepté. (mouille le doigt).</li></ul>

## 5.2. Géométrie de la main

Cette méthode consiste à déterminer les caractéristiques de la main d'un individu : sa forme, la longueur, la largeur, la courbure des doigts, etc. les systèmes de la géométrie de la main sont simples d'usage. L'utilisateur doit poser la paume de sa main sur une plaque qui possède des guides afin de l'aider à positionner ses doigts. Ces appareils peuvent être difficiles à utiliser pour certaines catégories de population pour lesquelles étendre la main est un problème, telles que les personnes âgées ou celles qui ont de l'arthrite. Une photo de face de la main est ensuite prise par un appareil photo numérique. Une photo de profil peut aussi être prise pour obtenir de l'information sur l'épaisseur de la main [6].



**FIG4** : dispositif de reconnaissance par géométrie de la main

Le tableau suivant résume les principales caractéristiques et les limites de la technique :

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• le résultat est indépendant de l'humidité des doigts de souillures éventuelles car il n'y a pas de contact direct avec le capteur ou une fenêtre, donc pas de risque d'encrassement.</li> <li>• Facilité de l'enrôlement du point de vue de l'utilisateur et bonne acceptation psychologique.</li> <li>• Faible volume de stockage par fichier.</li> </ul>	<ul style="list-style-type: none"> <li>• Système encombrant.</li> <li>• Risque élevé du taux de fausses acceptations et faux rejets, par exemple à cause d'une blessure ou pour les jumeaux ou les membres d'une même famille.</li> <li>• Cette technique n'a pas évolué depuis plusieurs années.</li> <li>• Le lecteur est plus cher que les autres types de capture de données physique.</li> </ul>

### 5.3. Visage

Le visage est certainement la caractéristique biométrique que les humains utilisent le plus naturellement pour s'identifier entre eux, ce qui peut expliquer pourquoi elle est en général très bien acceptée par les utilisateurs. Le système d'acquisition est soit un appareil photo, soit une caméra numérique [6].

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• Technique peu coûteuse, peu encombrante.</li> <li>• Absence de contact avec le capteur, méthode non intrusive pour la personne ; pas de risque pour la santé.</li> </ul>	<ul style="list-style-type: none"> <li>• Les vrais jumeaux ne sont pas différenciés.</li> <li>• En tant que contrôle d'accès, le visage n'est pas traditionnellement, reconnu comme un mécanisme fiable d'authentification. (peut être dupé par l'utilisation de maquillage ou d'un masque en silicone)</li> <li>• Tout élément tel que lunettes de soleil, chapeau, moustache, barbe, piercing, blessure peut causer des</li> </ul>

	<p>anomalies avec des systèmes d'identification du visage.</p> <ul style="list-style-type: none"><li>• Technique trop sensible au changement d'éclairage, d'échelle (taille du visage ou distance de la caméra), présence d'arrière-plan non stationnaire, changement de position lors de l'acquisition de l'image (inclinaison de la tête ou expression).</li><li>• Psychologiquement, certaines personnes rejettent leur image photographique (refus de son image, ajout d'accessoires, rôle, religion, critique de la qualité de la caméra, etc.). l'image est considérée comme trop personnelle pour être utilisée.</li></ul>
--	---

#### 5.4. La rétine

La reconnaissance de la rétine est une méthode assez ancienne puisque les premières études remontent aux années 30. On mesure dans cette technique la forme des vaisseaux sanguins qui se trouvent dans l'arrière de l'œil. Le dispositif utilisé emploie une source lumineuse rayonnante en direction de l'œil de l'utilisateur, qui se doit se tenir fixe devant le dispositif [6].



**FIG 5:** Photo de rétine

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• Résistant à la fraude, difficile et long à imiter.</li> <li>• Unicité même chez les vrais jumeaux</li> <li>• Technique fiable.</li> <li>• La cartographie de la rétine est la même tout au long de la vie, en absence de maladie spécifique.</li> </ul>	<ul style="list-style-type: none"> <li>• Nécessité de placer ses yeux à très faible distance du capteur, donc système intrusif mal accepté psychologiquement.</li> <li>• Difficile à utiliser en cas de contrôle d'une population importante (temps important).</li> </ul>

### 5.5. L'iris

La reconnaissance de l'iris est une technologie plus récente puisqu'elle s'est véritablement développée que dans les années 80. L'iris est la région annulaire située entre la pupille et le blanc de l'œil. Après l'avoir localisé, on prend des photos en noir et blanc, on utilise ensuite des coordonnées polaires et on cherche les transformées en ondelettes, pour avoir finalement un code représentatif de l'iris. Et on utilise la distance de hamming comme mesure de similarité [6].

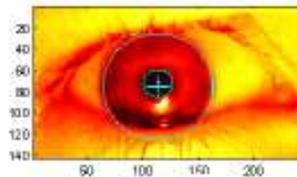


FIG 6 : Photo d'iris

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• Fiable.</li> <li>• Pas de risque identifié pour la santé.</li> </ul>	<ul style="list-style-type: none"> <li>• Système intrusif mal accepté psychologiquement.</li> <li>• Contraintes d'éclairage.</li> </ul>

### 5.6. Voix

La reconnaissance du locuteur vise à déterminer les caractéristiques uniques de la voix de chaque individu. Bien que généralement classée comme caractéristique

comportementale, la voix se trouve à la frontière avec les caractéristiques physiques. En effet, une grande partie de cette caractéristique est déterminée par le conduit vocal ainsi que les cavités buccales et nasales [6].



FIG 7 : Spectre d'un signal voix

Avantages	Inconvénients
<ul style="list-style-type: none"><li>• Disponible via le réseau téléphonique</li><li>• Les imitateurs utilisent les caractéristiques vocales sensibles au système auditif humain, mais ne sont pas capables de recréer les harmoniques de la voix, servant de base à l'identification. Il est quasi impossible d'imiter la voix stockée dans la base de données.</li><li>• Non intrusif</li></ul>	<ul style="list-style-type: none"><li>• L'utilisation d'un micro nécessite un dispositif adapté présent sur l'environnement.</li><li>• Sensibilité à l'état physique et émotionnel d'un individu.</li><li>• Sensibilité aux conditions d'enregistrement du signal de parole : bruit ambiant, parasites, qualité du microphone utilisé, qualité de l'équipement, lignes de transition.</li><li>• Fraude possible en utilisant un enregistrement de la voix de la personne autorisée, facilitée dans le cas de système basé sur la lecture d'un texte fixe.</li></ul>

### 5.7. Signature dynamique

La vérification par signature est une méthode automatique de mesure des signatures des personnes. Cette technologie examine un ensemble des dynamiques comme la vitesse, la direction, la pression de l'écriture, le temps pendant lequel le stylo est en contact avec le papier, le temps pris pour faire la signature et les positions où le stylo est relevé et abaissé sur le papier. La capture se fait à l'aide d'une tablette graphique [6].

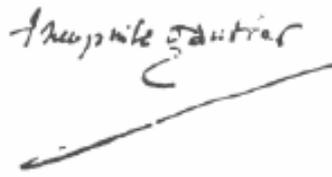


FIG 8 : Signature

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• Geste naturel qui responsabilise le signataire.</li> </ul>	<ul style="list-style-type: none"> <li>• Détermination d'un seuil.</li> <li>• Dépendance de l'état de santé ou émotionnel de la personne.</li> </ul>

### 5.8. Dynamique de la frappe au clavier

La dynamique de la frappe au clavier est caractéristique de l'individu, c'est en quelque sorte la transposition de la graphologie aux moyens électroniques. Les paramètres suivants sont généralement pris en compte :

- ✓ Vitesse de frappe.
- ✓ Suite de lettres.
- ✓ Mesure des temps de frappe.
- ✓ Pause entre chaque mot.
- ✓ Reconnaissance de mots précis.

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• Moyen non intrusif qui exploite un geste naturel.</li> </ul>	<ul style="list-style-type: none"> <li>• Dépendance de l'état physique de la personne. (âge, maladies,...)</li> </ul>

L'utilisation d'une seule caractéristique biométrique pour l'identification de l'utilisateur ne peut pas garantir de bonnes performances dans certains cas, par exemple l'identification de visage dans le cas de vrais jumeaux, la variation du signal acquis ou la présence de bruit. Afin de surmonter ce problème, des méthodes biométriques appelées multimodales [9], ont été introduites pour obtenir de meilleures performances. Il s'agit de combiner plusieurs biométries dans le processus de l'identification pour assurer un niveau de sécurité plus élevé. Un exemple de ces méthodes est celle qui utilise le visage, la voix et le mouvement des lèvres ensemble.

## 6. Evaluation d'un système biométrique

Devant cette diversité des techniques biométriques, l'analyse comparative des performances des différentes techniques biométriques est importante avant la prise de décision d'implantation d'un système, et cela pour décider si la biométrie est adaptée à la situation et aux fonctions recherchées et dans un deuxième temps de décider quelle technique biométrique est adaptée et donne les meilleurs résultats. Plusieurs études de ce type ont été menées. Elles essayaient de trouver les critères d'évaluation et de comparaison. Ainsi l'International Biometric Group, une société américaine d'intégration et de conseil, a effectué une étude basée sur quatre critères d'évaluation [10]:

- **L'effort** : requis par l'utilisateur lors de la prise de mesures biométriques, et qui doit être réduit le plus possible. La reconnaissance faciale est la technique biométrique la plus facile à utiliser car non contraignante.
- **L'intrusivité** : ce critère permet de classifier les systèmes biométriques en fonction de l'existence d'un contact direct entre le capteur utilisé et l'individu à reconnaître. La reconnaissance faciale est une technique « non intrusive », car il n'existe aucun contact entre le capteur (la caméra) et le sujet, elle est bien acceptée par les utilisateurs à l'inverse d'autres techniques « intrusives » comme l'iris où un contact direct est nécessaire entre le capteur et l'œil.
- **Le coût** : doit être modéré. À cet égard nous pouvons dire que la reconnaissance faciale ne nécessite pas une technologie coûteuse. En effet, la plupart des systèmes fonctionnent en utilisant un appareil à photo numérique de qualité standard.
- **La fiabilité** : dépend de la qualité de l'environnement (éclairage par exemple) dans lequel l'utilisateur se trouve. Ce critère influe sur la reconnaissance de l'utilisateur par le système. La fiabilité de la reconnaissance faciale est controversée. Elle est considérée comme une technologie relativement peu sûre, car le signal acquis est sujet à des variations beaucoup plus élevées que pour d'autres caractéristiques morphologiques. Ces variations peuvent être causées, entre autres, par le maquillage, la pilosité, la présence ou l'absence de lunettes, le vieillissement et l'expression d'une émotion. La méthode d'authentification de visage est sensible à la variation de l'éclairage et au changement de position du visage lors de l'acquisition de l'image.

Le schéma suivant représente les différentes techniques biométriques, évaluées selon les critères établis par l'International Biometric Group [10].

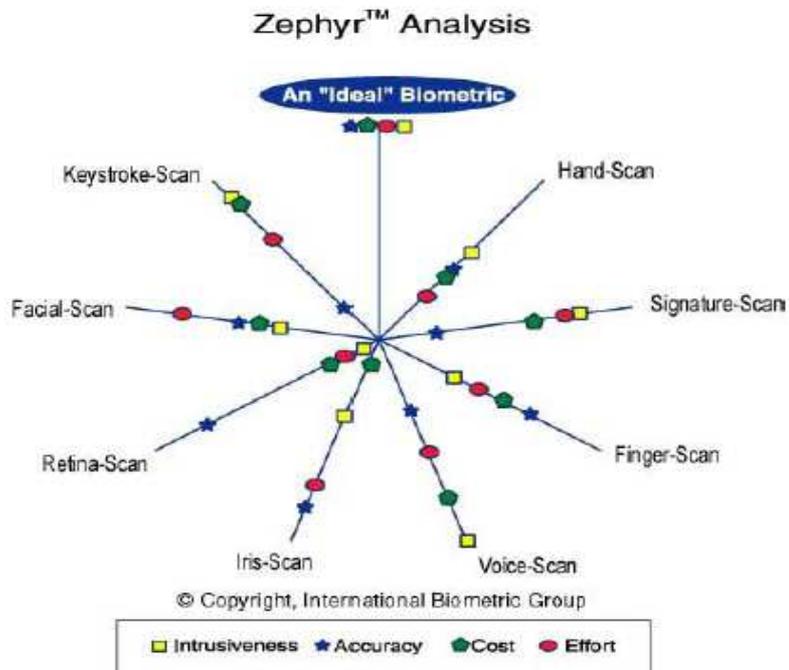


FIG 9 : les différents critères pour chaque type de technique biométrique [10]

**Légende :**

- **Effort** : effort requis par l'utilisateur lors de la mesure.
- **Intrusiveness**: décrit dans quelle mesure l'utilisateur perçoit le test comme intrusif.
- **Cost**: coût de la technologie (lecteur, capteur, etc....)
- **Accuracy**: efficacité de la méthode (capacité à identifier quelqu'un)

Les résultats de cette étude est comme on le voit résumer sur la figure 9, peuvent être interprétés comme suit :

- Les techniques les moins "intrusives" aux plus "intrusives" : la voix, la frappe sur le clavier, la signature, la main, le visage, l'empreinte digitale, l'iris et enfin la rétine.
- Les techniques les plus fiables aux moins fiables: l'iris, la rétine, l'empreinte digitale, le visage, la main, la voix, et enfin à un niveau équivalent, la frappe sur le clavier et la signature.
- Les techniques les moins coûteuses aux plus coûteuses: la frappe sur le clavier, la voix, la signature, l'empreinte digitale, le visage, la main, la rétine et enfin l'iris.

- Les techniques les plus faciles d'utilisation aux plus difficiles: le visage, la signature, l'iris, la frappe sur le clavier, la voix, l'empreinte digitale, la main et enfin la rétine.

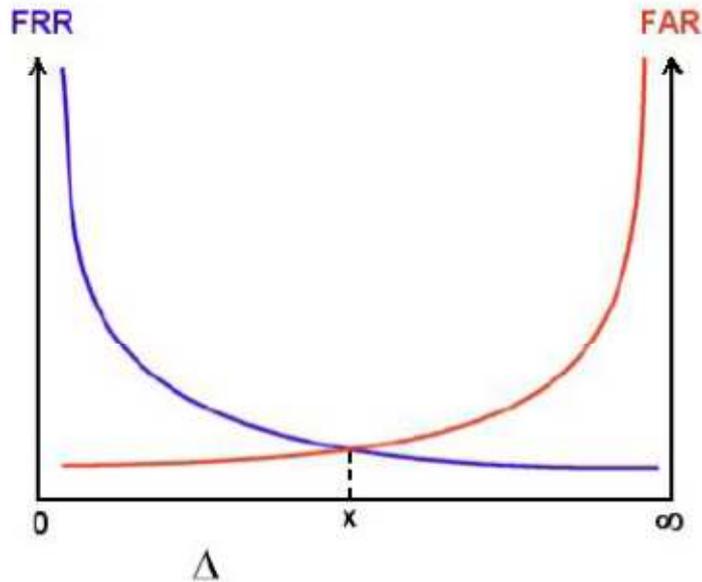
## 7. La fiabilité des systèmes biométriques

La fiabilité des systèmes biométriques est liée à deux facteurs: l'unicité des données biométriques et les taux d'erreurs.

La question de l'unicité est importante puisque les systèmes biométriques prétendent relier une donnée à une personne. Elle recouvre en fait deux problèmes: l'unicité du caractère biométrique choisi et l'unicité de la mesure ou de la représentation graphique de ce caractère. L'unicité des caractères biométriques n'est pas prouvée scientifiquement. La biométrie se repose alors sur des méthodes statistiques destinées à déterminer la probabilité que deux personnes présentent la même donnée. C'est le cas des empreintes digitales dont on possède des bases de données suffisamment larges (400 millions d'ID dans les Fichiers du FBI) et dans lesquelles n'ont jamais été constatées deux empreintes identiques. En revanche aucun test n'aurait été réalisé pour déterminer la probabilité que deux personnes puissent avoir le même visage. Par ailleurs même si les caractéristiques humaines prises en compte étaient uniques, les techniques appliquées et les différentes circonstances dans lesquelles les caractéristiques humaines sont présentées et mesurées jouent un rôle déterminant. Ainsi pour dix points de comparaison, la probabilité de trouver les mêmes points disposés de façon identique sur les empreintes digitales de deux personnes différentes serait d'une chance sur un million et, pour 14 à 17 points, d'une chance sur 17 milliards. D'une façon similaire, certaines études proposent que 9 points bien choisis sont suffisants pour identifier un visage parmi un groupe de plus de 80 personnes mais que pour des bases de données plus larges le nombre de points caractéristiques doit être plus grand [11]. Deux notions sont importantes pour évaluer un système biométrique :

- **le taux de fausses acceptations (FAR)** qui est le ratio entre le nombre de personnes qui ont été acceptées alors qu'elles n'auraient pas dû l'être et le nombre total de personnes non autorisées qui ont tenté de se faire accepter.
- **Le taux de faux rejets (FRR)** qui est le ratio entre le nombre de personnes autorisées dont l'accès a été refusé et le nombre total de personnes autorisées s'étant présenté.

Les deux types de taux sont reliés de façon inverse, l'équilibre entre ces deux taux qui permet d'obtenir les meilleurs résultats doit être adapté en fonction des besoins (taux de fausse acceptation très diminuée pour un système nécessitant une haute sécurité avec conséquence inévitable un taux de faux rejet élevé). La figure 7 nous montre cet équilibre.



**FIG 10:** La recherche d'un compromis entre les 2 taux FRR et FAR résulte le point X, la zone delta représente la marge d'erreur autorisée par le système.

## 8. Les applications de la biométrie

Les applications de biométrie peuvent être divisées en quatre groupes principaux :

### 8.1. Contrôle d'accès

Le contrôle d'accès peut être lui-même subdivisé en deux sous catégories : le contrôle d'accès physique et le contrôle d'accès virtuel. On parle de contrôle d'accès physique lorsqu'un utilisateur cherche à accéder à un lieu sécurisé. On parle de contrôle d'accès virtuel dans le cas où un utilisateur cherche à accéder à une ressource ou un service.

#### 8.1.1. Contrôle d'accès physique

Longtemps, l'accès à des lieux sécurisés (bâtiments ou salles par exemple) s'est fait à l'aide de clefs ou badges. Les badges étaient munis d'une photo et un garde chargé de la

vérification. Grâce à la biométrie, la même opération peut être effectuée automatiquement de nos jours.

L'une des utilisations les plus célèbres de la géométrie de la main pour le contrôle d'accès est le système INSPASS [12] (Immigration and Naturalization Service Passenger Accelerated Service System) déployé dans plusieurs grands aéroports américains (New York, Washington, Los Angeles, San Francisco, etc.) cette application permet aux passagers répertoriés dans le système d'éviter les files d'attente pour le contrôle des passeports. Ceux-ci possèdent une carte magnétique qui contient l'information sur la géométrie de leur main. Lorsqu'ils présentent leur main au système, celle-ci est comparée à l'information contenue dans la carte.

### **8.1.2. Contrôle d'accès virtuel**

Le contrôle d'accès virtuel permet par exemple l'accès aux réseaux d'ordinateurs ou l'accès sécurisé aux réseaux d'ordinateurs ou l'accès aux sites web. Le marché du contrôle d'accès virtuel est dominé par les systèmes basés sur une connaissance, typiquement un mot de passe. Avec la chute des prix des systèmes d'acquisitions, les applications biométriques devraient connaître une popularité croissante.

Un exemple d'application est l'intégration par Apple dans son système d'exploitation MAC OS 9 d'un module de reconnaissance du locuteur [13] de manière à protéger les fichiers d'un utilisateur, tout particulièrement lorsque l'ordinateur est utilisé par plusieurs individus ce qui est de plus en plus souvent le cas.

## **8.2. Authentification des transactions**

L'authentification des transactions représente un marché gigantesque puisqu'il englobe aussi bien le retrait d'argent au guichet des banques, les paiements par cartes bancaires, les transferts de fond, les paiements effectués à distance par téléphone ou sur Internet, etc.

Mastercard estime ainsi que les utilisations frauduleuses de cartes de crédit pourraient être réduites de 80% en utilisant des cartes à puce qui incorporeraient la connaissance des empreintes digitales [14]. Les 20% restant seraient principalement dus aux paiements à distance pour les quelles il existerait toujours un risque. Pour les transactions à distance, des solutions sont déjà déployées en particulier pour les

transactions par téléphone. Ainsi, la technologie de reconnaissance du locuteur de Nuance (Nuance Verifier™ [Nua]) est utilisée par les clients du Home Shopping Network, une entreprise de téléshopping.

### **8.3. Répression**

Une des applications les plus immédiates de la biométrie à la répression est la criminologie. La reconnaissance des empreintes digitales en est l'exemple le plus connu. Elle fut acceptée dès le début du xx<sup>e</sup> siècle comme moyen d'identifier formellement un individu et son utilisation s'est rapidement répandue.

Il existe aussi des applications dans le domaine judiciaire. T-Netix [15] propose ainsi des solutions pour le suivi des individus en liberté surveillée en combinant technologie de l'Internet et de reconnaissance de locuteur.

### **8.4. Personnalisation**

Les technologies biométriques peuvent être aussi utilisées afin de personnaliser les appareils que nous utilisons tous les jours. Cette application de la biométrie apporte un plus grand confort d'utilisation.

Afin de personnaliser les réglages de sa voiture, Siemens propose par exemple d'utiliser la reconnaissance des empreintes digitales [16]. Une fois l'utilisateur identifié, la voiture ajuste automatiquement les sièges, les rétroviseurs, la climatisation, etc.

## **9. La place de la reconnaissance faciale parmi les autres techniques biométriques**

Les empreintes digitales sont les caractéristiques biométriques les plus communément utilisées pour la reconnaissance de criminels. Le premier système automatique d'authentification utilisant les empreintes digitales a été commercialisé au début des années soixante. D'autres parts, plusieurs études ont démontré que l'iris est la caractéristique la plus fiable car la texture de l'iris reste stable au cours de la vie. Toutefois, ces méthodes présentent l'inconvénient majeur d'être intrusives, ce qui limite énormément leurs domaines d'applications. De plus, une méthode comme l'identification de l'iris reste contraignante pour les utilisateurs qui n'apprécient pas de placer leur œil

devant un appareil. A l'inverse, des systèmes d'identification basés sur l'analyse des images de visage ne présentent aucune contrainte pour les utilisateurs. La reconnaissance faciale est une modalité qui peut être implémentée de manière indépendante des autres modalités biométriques, elle est souvent utilisée dans des applications de surveillance. La reconnaissance faciale offre plusieurs avantages : le système de capture (c.à.d. les caméras) est facile à installer, il est accepté dans les lieux publics ce qui permet d'avoir des bases de données de plus en plus grandes et ainsi d'améliorer les performances de la reconnaissance.

Depuis quelques années, la reconnaissance faciale suscite un intérêt croissant auprès de la communauté scientifique, qui s'est notamment manifesté à travers l'organisation de conférences internationales spécialisées telle que « The International Conference on Audio and Video-based Biometric Person Authentication (AVBPA) » depuis 1997, et « the International Conference on Automatic Face and Gesture Recognition (AFGR) » depuis 1995. Par ailleurs, les États-Unis ont mené depuis 1993 une série de tests de reconnaissance faciale dont les résultats sont accessibles au public. Ces tests sont désignés sous les noms de FERET [17][10][18], XM2VTS [19], FRVT 2000 [20] et FRVT 2002 [21]. Dans [22] les auteurs présentent une analyse statistique des publications sur les techniques biométriques soumises et publiées dans un numéro spécial de la revue *IEEE Transaction on PAMI* (voir tableau 1). Nous constatons que la reconnaissance faciale arrive largement en tête avec un pourcentage de 33% du nombre total de publications. Ceci démontre bien l'intérêt scientifique pour cette technique.

Article (%)	Visage	Empreintes digitale	Multimodale	Iris	Performance Evaluation	Autres
Soumission	33%	17%	16%	9%	4%	21%
Acceptation	33%	16%	16%	11%	5%	20%

**Tableau 1:** Répartition des articles sur les techniques biométriques soumis et acceptés dans la revue IEEE PAMI [22]

## 10. Conclusion

La biométrie offre beaucoup plus d'avantages que les méthodes existantes d'authentification personnelle, elle fournit plus de sûreté et de convenance. Il existe plusieurs techniques biométriques utilisées tel que: l'iris, la rétine, et l'empreinte digitale,

mais ces techniques présentent l'inconvénient d'être intrusives (au sens où elles nécessitent la coopération de l'utilisateur) ce qui réduit le champ des applications. De plus, leur déploiement repose sur l'utilisation d'un matériel dédié. A l'inverse, la technique de reconnaissance de visages est non-intrusive (on peut vérifier l'identité de quelqu'un sans même qu'il s'en rende compte), et un matériel de prise de vue courant (comme par exemple un appareil photographique numérique ou une webcam) suffit pour l'acquisition des données.

## Chapitre II

# La reconnaissance des visages humains

### 1. Introduction

La reconnaissance des visages constitue l'une des capacités les plus étonnantes, et les plus performantes, du système visuel humain. En effet, l'être humain est capable de discriminer et de reconnaître un très grand nombre de visages, alors même que les visages constituent une catégorie de stimulus homogènes, partageant un ensemble de traits similaires et une structure commune, et qu'ils peuvent subir d'importants changements avec l'âge par exemple, ou encore selon l'expression faciale. Malgré les efforts fournis par les chercheurs afin d'automatiser la reconnaissance des visages, ils ne sont pas encore parvenus à un système totalement fiable à 100%, car la machine est très loin de pouvoir simuler le cerveau humain. Cependant beaucoup de chercheurs essaient d'automatiser le processus de reconnaissance des visages. Pour cela différentes théories mathématiques et statistiques trouvent leurs applications dans le domaine de la reconnaissance des visages. Il faut adapter ces méthodes à ce problème, en essayant de lui trouver un modèle représentatif. Ces méthodes ont été implémentées et ont donné des résultats intéressants et satisfaisants mais aucune d'elle n'a atteint l'exactitude en raison des multiples paramètres qu'il faut prendre en considération tel que : la variation de posture, l'éclairage, le style de coiffure, le port de lunettes, la barbe, les moustaches, la vieillesse, etc. en plus de la complexité de la représentation du visage humain. Il est très important de s'intéresser à ce domaine vu ses nombreux champs d'applications tel que : la vérification d'identité, la télésurveillance, les interfaces homme machines, etc.

## 2. Motivations : (pourquoi la reconnaissance de visages ?)

Durant les vingt dernières années, la reconnaissance automatique des visages est devenue un enjeu primordial, ceci est dû à ses caractéristiques avantageuses dont on peut citer :

- La disponibilité des équipements d'acquisition, leur simplicité et leurs coûts faibles.
- Passivité du système : un système de reconnaissance de visages ne nécessite aucune coopération de l'individu, du genre : mettre le doigt ou la main sur un dispositif spécifique ou parler dans un microphone. En effet, la personne n'a qu'à rester ou marcher devant une caméra pour qu'elle puisse être identifiée par le système.

En plus, cette technique est très efficace pour les situations non standards, c'est les cas où on ne peut avoir la coopération de l'individu à identifier, par exemple lors d'une arrestation des criminels. Certes que la reconnaissance des visages n'est pas la plus fiable comparée aux autres techniques de biométrie, mais elle peut être ainsi si on utilise des approches plus efficaces en plus du bon choix des caractéristiques d'identification représentant le visage en question.

Le grand intérêt accordé à la reconnaissance des visages est dû à l'importance et à l'utilité du visage par rapport aux autres parties du corps humain. En effet à partir de l'image du visage d'une personne on peut deviner son identité, sa race, son sexe, etc. Cet intérêt s'explique aussi par la multitude et la variété des domaines d'applications possibles, parmi lesquelles on trouve :

- Les systèmes de contrôle d'accès automatique : Une caméra placée à l'entrée d'un site envoie les images des individus, désirant accéder à ce site, à un ordinateur qui effectue une reconnaissance des visages pour décider s'il faut accorder l'accès ou non.
- Les systèmes de télésurveillance et d'identification criminelle : Pour les enquêteurs la vérification des identités des suspects peut être assistée par ordinateur (c'est le cas du système WHQ\_IS) ou entièrement automatisée.
- Les interfaces homme-machine : On peut rendre la communication entre l'homme et la machine plus attractive, plus naturelle et plus confortable. Par exemple, s'il est possible d'estimer la direction de la tête, on peut diriger une caméra vers l'objet que l'individu regarde et acquérir des informations concernant cet objet. La téléconférence.
- Les langages des signes.
- ...

### 3. Qu'est ce qu'un système de reconnaissance de visage ?

Un système automatique de reconnaissance de visages est un système biométrique utilisant le visage à des fins d'identification et/ou de vérification de personnes à partir de leurs images de visages fixes[23] ou de séquences vidéo[24] en comparant les caractéristiques de cet individu avec celles stockées dans une base de référence.

Tout processus automatique de reconnaissance de visages doit prendre en compte plusieurs facteurs qui contribuent à la complexité de sa tâche, car le visage est une entité dynamique qui change constamment sous l'influence de plusieurs facteurs. La figure 11, illustre la démarche générale adoptée pour réaliser de tels systèmes [25] :

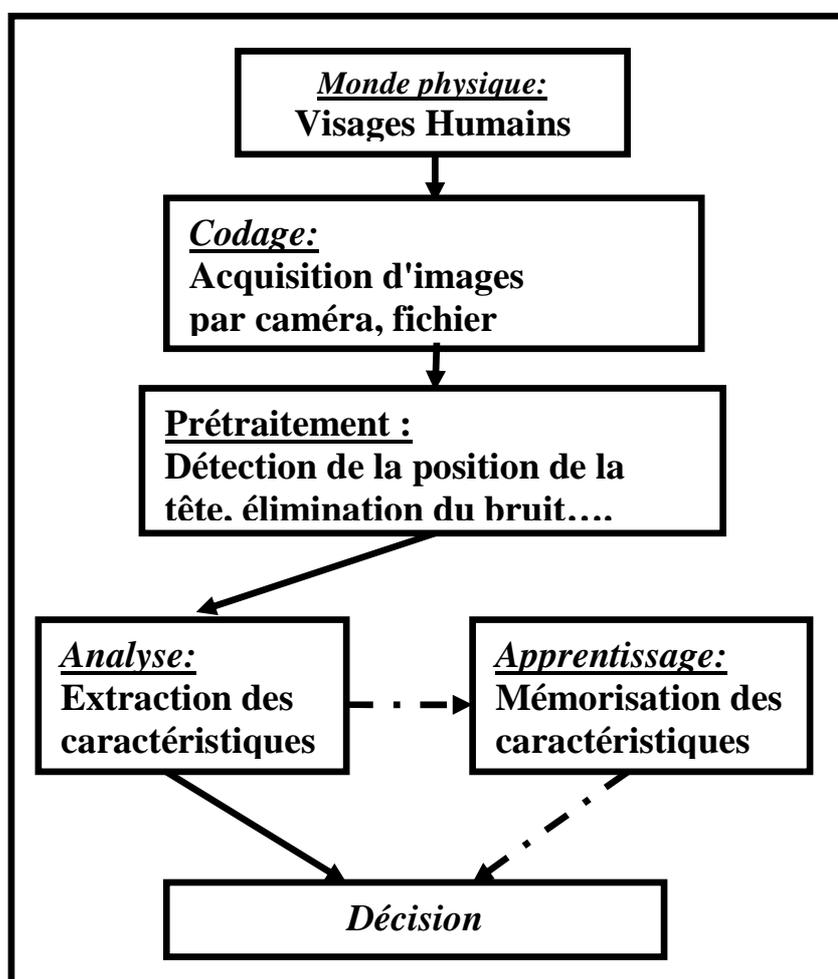


FIG 11: Schéma général d'un système de reconnaissance de visages[25]

### **3.1. Le monde physique :**

Dans l'environnement où se trouve l'individu à identifier, il y a trois paramètres à considérer :

- L'éclairage : L'image du visage est une matrice de niveaux de gris (C'est le cas dans la majorité des systèmes existants) représentant des intensités de lumière.
- La variation de posture qui dépend de la position de la tête et de l'angle de prise de vue (vue de face, de profile, 45° ...).
- L'échelle : qui dépend de la taille de l'image et de la distance séparant l'individu de la caméra, une normalisation des distances est donc nécessaire.

La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents.

### **3.2. Acquisition de l'image**

C'est l'opération qui permet d'extraire du monde réel une représentation bidimensionnelle pour des objets en 3D, cette opération peut être statique (appareil photo, Scanner...etc.) ou dynamique (Caméra, Webcam), dans ce cas on aura une séquence vidéo. A ce niveau on aura une image brute.

### **3.3. Prétraitement**

Les données brutes issues des capteurs sont des représentations initiales des données, à partir desquelles des traitements permettent de construire celles qui seront utilisées pour la reconnaissance. L'image brute peut être affectée par différents facteurs causant ainsi sa détérioration, elle peut être bruitée, c'est-à-dire contenir des informations parasites à cause des dispositifs optiques ou électroniques. Pour pallier à ces problèmes, il existe plusieurs méthodes de traitement et d'amélioration des images, telle que : la normalisation, l'égalisation de l'histogramme, etc.

### **3.4. Extraction de caractéristiques**

Dans cette étape on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision. Le choix de ces

informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes. Ces informations seront ensuite classées, en d'autres termes, elles seront affectées à la classe la plus proche, les individus ayant des similarités sont regroupés dans la même classe. Ces classes varient selon le type de décision.

### 3.5. Apprentissage

L'apprentissage consiste à mémoriser les modèles calculées dans la phase analyse pour les individus connus. Un modèle est une représentation compacte des images qui permet de faciliter la phase de reconnaissance mais aussi de diminuer la quantité de données à stocker en quelque sorte l'apprentissage est la mémoire du système.

### 3.6. Décision

La décision est la partie du système où on tranche sur l'appartenance d'un à l'ensemble des visages ou pas, et si oui quelle est son identité. Donc la décision c'est l'aboutissement du processus. On peut le valoriser par taux de reconnaissance (fiabilité) qui est déterminé par le taux de justesse de la décision.

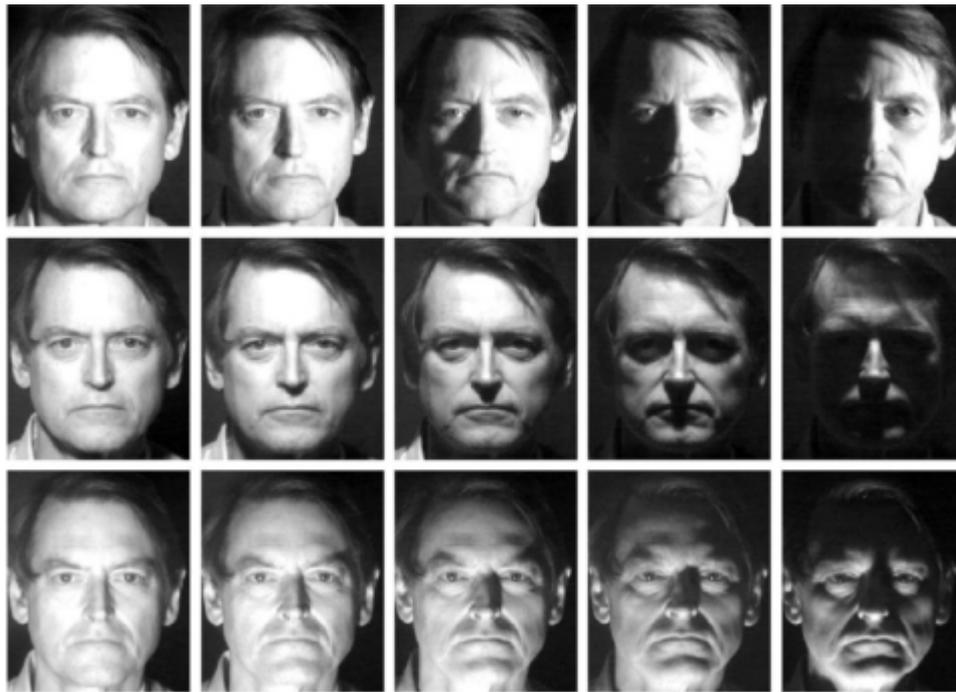
## 4. Difficultés inhérentes à la reconnaissance automatique de visage

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables.

### 4.1. Changement d'illumination

L'intensité et la direction d'illumination lors de la prise de vue influent toutes deux énormément sur l'apparence du visage dans l'image. Dans la plupart des applications réelles, des changements dans les conditions d'illumination sont néanmoins inévitables, notamment lorsque les vues sont collectées à des dates différentes, en intérieur ou en extérieur. Étant donné qu'un visage humain est un objet intrinsèquement 3D, des changements d'illumination peuvent faire apparaître sur le visage des ombres accentuant ou,

au contraire, masquant certaines caractéristiques faciales (voir figure 12). Ceci a été expérimentalement observé dans Adini et al [26] où les auteurs ont utilisé une base de données de 25 individus. L'identification de visage dans un environnement non contrôlé reste donc un domaine de recherche ouvert. Les évaluations FRVT [27] ont révélé que le problème de variation d'illumination constitue un défi majeur pour la reconnaissance faciale.



**FIG 12:** Extrait de [28]. Effets de variations dans les directions d'illumination sur l'apparence d'un visage

## 4.2. Les variations de pose

Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. Cette difficulté a été démontrée par des tests d'évaluation élaborés sur les bases FERET et FRVT [27,29]. La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Quand le visage est de profil dans le plan image (orientation  $< 30^\circ$ ), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à  $30^\circ$ , la normalisation géométrique n'est plus possible (voir figure 13).



FIG 13: Exemple de variation de pose

### 4.3. Occultations partielles

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoire tels que lunettes, écharpe... Dans le contexte de la biométrie, les systèmes proposés doivent être non intrusifs c'est-à-dire qu'on ne doit pas compter sur une coopération active du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés. Gross et al [30] ont étudié l'impact du port de lunettes de soleil, et du cache-nez occultant la partie inférieure du visage sur la reconnaissance faciale. Ils ont utilisé la base de données AR [31]. Leurs résultats expérimentaux semblent indiquer que, dans ces conditions, les performances des algorithmes de reconnaissance restent faibles.

### 4.4. Expressions faciales

Un autre facteur qui affecte l'apparence du visage est l'expression faciale (voir figure 14). La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu. L'information temporelle fournit une connaissance additionnelle significative qui peut être utilisée pour résoudre ce problème [30].



FIG 14: Exemple de variation d'expression

#### **4.5. Présence ou absence des composants structurels**

La présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance. Par exemple, des lunettes opaques ne permettent pas de bien distinguer la forme et la couleur des yeux, et une moustache ou une barbe modifie la forme du visage.

#### **4.6. L'impact de la taille de la base**

C'est lors du FRVT 2002 que l'impact de la taille de la base sur les performances du système a été étudié pour la première fois, grâce à la très grande taille de la base HCINT utilisée [32]. Les meilleurs systèmes fournissent des taux de reconnaissance de l'ordre de 85% pour 800 personnes, 83% pour 1600 individus, et 73% pour 37437 personnes. Selon les conclusions de ce rapport, les performances décroissent de manière log-linéaire en fonction de la taille de la base.

### **5. Les méthodes de reconnaissance de visage**

Les systèmes de reconnaissance de visage peuvent grossièrement être classés en trois grandes catégories : les méthodes locales « Feature-Based », les méthodes dites globales « Image-Based », et les méthodes hybrides.

#### **5.1. Les méthodes locales**

Ce sont des méthodes géométriques, on les appelle aussi les méthodes à traits, à caractéristiques locales, ou analytiques. L'analyse du visage humain est donnée par la description individuelle de ses parties et de leurs relations. Ce modèle correspond à la manière avec laquelle l'être humain perçoit le visage, c'est-à-dire, à nos notions de traits de visage et de parties comme les yeux, le nez, la bouche, etc. Elles consistent à localiser en premier lieu les éléments caractéristiques du visage tel que les yeux, le nez, la bouche, les joues... et utiliser, pour la reconnaissance, soit les régions contenant ces éléments [33] soit les relations géométriques entre eux (distance entre les deux yeux, entre les deux joues etc.), à partir des vues de profil [34] ou de face [35, 36, 37].

L'avantage de ces méthodes est qu'elles prennent en compte la particularité du visage en temps que forme naturelle à reconnaître, en exploitant les résultats de la recherche en neuropsychologie et psychologie cognitive concernant le système visuel humain, ainsi qu'un nombre réduit de paramètre (de 9 à 14 distances au maximum à considérer) [38, 25].

Les capacités de ces méthodes sont limitées. Cela est dû essentiellement à l'occultation de certaines caractéristiques (vue de profil, port de lunettes, moustache, cache-nez....) et aux algorithmes de localisation des éléments caractéristiques de visages qui sont jusqu'à présent peu performants.

### 5.2. Les méthodes globales

Contrairement aux méthodes géométriques, les méthodes globales ne nécessitent pas des connaissances préalables sur l'objet caractéristique à localiser dans le visage. Elles utilisent l'information apportée par le visage entier sans segmentation de ses parties. Elles se basent principalement sur l'information pixel. Pour les appliquer, les visages de la base de données sont représentés par des matrices de  $n$  lignes et de  $m$  colonnes chacun, doivent être transformés en vecteurs en concaténant les lignes (ou les colonnes) de ces matrices afin d'obtenir des vecteurs de dimensions  $(n \times m, 1)$  qui seront fournis au classificateur.

La force des méthodes globales tient à ce qu'elles utilisent la totalité des caractéristiques du visage, en ne réservant pas un traitement préférentiel à certaines « régions ». En général, les méthodes globales fournissent de bons taux de reconnaissance, mais nécessitent que le visage soit présenté dans un cadre simple : visage présenté à peu près de face, éclairage régulier, arrière-plan simple.

Les performances se dégradent rapidement dès qu'il y a des changements d'orientation du visage, que l'éclairage varie brusquement ou que l'arrière-plan est trop chargé. Le problème de stockage des informations extraites lors de la phase d'apprentissage reste leur l'inconvénient majeur [39, 40].

### 5.3. Les méthodes hybrides

Les psychophysiciens et neuroscientifiques ont cherché à déterminer si le cerveau humain se base plutôt sur l'étude de caractéristiques globales ou locales du visage pour sa reconnaissance. De nos jours, la plupart des chercheurs s'accordent à dire que les deux

types d'information sont utilisées [41]. Alors la robustesse d'un système de reconnaissance peut être augmentée par la fusion de plusieurs méthodes. Cette technique de reconnaissance consiste à combiner plusieurs méthodes pour résoudre le problème d'identification. Le choix d'un outil de reconnaissance robuste n'est pas une tâche triviale, cependant l'utilisation d'une approche multi-classifieur pour l'identification des visages est une solution adéquate à ce problème [42].

### 6. Les algorithmes classiques de reconnaissance de visages

Un grand nombre d'algorithmes appliqués à la reconnaissance du visage, la citation de la totalité des algorithmes ne peut être complète. On va décrire la plupart algorithmes classiques approches qui ont apporté une avancée sensible dans le domaine de reconnaissance de visage.

#### 6.1. Analyse en composantes principales

L'analyse en composantes principales PCA [43, 44] est une méthode mathématique qui peut être utilisé pour simplifier un ensemble de données, en réduisant sa dimension. Elle sert à construire un sous-espace plus approprié pour représenter les images exemples. La méthode cherche les axes orthogonaux pour lesquels la variance des images exemples projetées est maximale. Ces axes conviennent particulièrement bien pour représenter les images exemples, mais ils ne permettent par contre pas forcément de bien séparer les images exemples si celles-ci appartiennent à plusieurs classes. L'axe possédant la variance maximale est appelée premier axe principal. Dans tout l'espace des images, il n'existe aucun vecteur de projection avec une variance plus grande que celle du premier axe principal. Le deuxième axe principal est l'axe qui est perpendiculaire au premier axe principal et qui pointe dans la direction pour laquelle la projection des images exemples possède de nouveau la variance maximale. Les axes principaux suivants possèdent les mêmes propriétés par rapport aux axes principaux précédents que le deuxième axe principal par rapport au premier. Comme les premiers axes principaux possèdent les plus grandes variances, ce sont eux qui permettent de représenter le mieux les images exemples. La procédure pour calculer les axes principaux normalisés  $u_i$  est la suivante. D'abord, il faut calculer l'image moyenne des images exemples :

$$\bar{\mathbf{X}} = \frac{1}{M} \sum_{i=1}^M \mathbf{X}_i$$

Ensuite, il faut soustraire à chaque image exemple l'image moyenne :

$$\hat{\mathbf{X}}_i = \mathbf{X}_i - \bar{\mathbf{X}}, \text{ pour } i = 1, \dots, M$$

Les vecteurs  $\hat{\mathbf{X}}_i$  sont rassemblés dans la matrice

$$\mathbf{D} = [\hat{\mathbf{X}}_1 \hat{\mathbf{X}}_2 \dots \hat{\mathbf{X}}_M]$$

et la matrice de covariance

$$\mathbf{C} = \frac{1}{M-1} \mathbf{D} \mathbf{D}^T$$

est calculée. Maintenant, les vecteurs propres  $\mathbf{v}_i$  avec leur valeur propre  $\lambda_i$  sont calculés et mis dans l'ensemble

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M\}$$

de façon à ce que

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$$

Le vecteur propre  $\mathbf{v}_i$  de la matrice de covariance  $\mathbf{C}$  pointent dans la même direction que la  $i$ -ème axe principal normalisé  $\mathbf{u}_i$  de  $\mathbf{D}$ . On peut alors calculer les axes principaux  $\mathbf{u}_i$  :

$$\mathbf{u}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} \text{ pour } i = 1, 2, \dots, M$$

Une image  $\mathbf{X}$  peut être projetée dans le sous-espace en effectuant le calcul

$$\boldsymbol{\omega} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \cdot \\ \mathbf{u}_Q^T \end{bmatrix} \cdot \mathbf{X}$$

avec  $\boldsymbol{\omega}$  le vecteur des poids et  $Q$  nombre d'axes principaux choisis. L'image  $\mathbf{X}$  peut être reconstruite approximativement par

$$\mathbf{X}_r \approx [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdot \quad \mathbf{u}_Q] \cdot \boldsymbol{\omega}$$

avec une erreur de reconstruction

$$\varepsilon = \|\mathbf{X} - \mathbf{X}_r\|^2$$

L'erreur de reconstruction  $\varepsilon$  donne une indication de la distance entre l'image  $\mathbf{X}$  et le sous espace. Si les images exemples comprennent seulement des images positives, alors l'erreur de reconstruction  $\varepsilon$  peut être directement utilisée comme classificateur. On supposera que l'objet caractéristique à localiser se trouve dans la région avec l'erreur de reconstruction  $\varepsilon$  minimale dans l'image scannée. Mais si les images exemples consistent d'images positives et négatives, alors un classificateur statistique sera nécessaire.

## 6.2. Les réseaux de neurones

D'après [45], Les réseaux de neurones artificiels sont des assemblages fortement connectés d'unités de calcul. Chacune des unités de calcul est un neurone formel qui est, en soi, une formulation mathématique d'un neurone biologique. Les réseaux de neurones artificiels sont dotés de capacités de mémorisation et d'une certaine forme d'apprentissage. Pour débiter, une image brute (ou prétraitée) de dimensions fixes constitue habituellement la source d'entrée des réseaux. Les dimensions doivent être établies au préalable car le nombre de neurones sur la couche d'entrée en dépend. Cela étant dit, plus les dimensions de l'image sont élevées, plus la complexité et le temps d'apprentissage augmentent. Un réseau de neurones [46] peut être un perceptron multicouche. L'unité de base du réseau de neurones est le perceptron. Chaque perceptron effectue un travail relativement simple : il reçoit des données pondérées des voisins ou des sources externes et calcule sur cette base un signal de sortie qui est propagé à d'autres unités. On distingue entre unité d'entrée, de sortie et cachée. Un réseau de neurones doit être configuré pour que l'application d'un ensemble de données d'entrées produise le résultat désiré à la sortie. Il y a plusieurs façons de configurer un réseau de neurones :

- Une est de choisir les poids explicites en utilisant du savoir à priori.
- Une autre est d'entraîner le réseau de neurones en lui donnant des images exemples et en le laissant changer la valeur de ses poids selon une règle d'apprentissage.

Pour la deuxième façon de configuration, il existe deux situations d'apprentissage distinctes :

- L'apprentissage supervisé. Dans ce cas, les images exemples doivent être classifiées.
- L'apprentissage non-supervisé. Dans ce cas, les images exemples ne doivent pas être classifiées. Le système est supposé découvrir les images qui se ressemblent statistiquement et de les attribuer à une classe.

### 6.3. Les machines à vecteur de supports

Une machine à vecteurs de support est une technique de discrimination, c'est une méthode d'apprentissage supervisé utilisée pour la classification et la régression. Elle consiste à séparer deux ou plusieurs ensembles de points par un hyperplan [47]. L'hyperplan est appelé le « meilleur » lorsqu'il maximise la marge entre lui et les deux classes. Cela veut dire qu'il maximise la distance entre lui et les points les plus proches des deux classes. Pour construire l'hyperplan, l'algorithme cherche à minimiser la fonction coût

$$F(\mathbf{w}) = \mathbf{w}^T \mathbf{w}$$

en respectant la contrainte

$$C_1 : y_i (\mathbf{w}^T \cdot \mathbf{x}_i + \omega_0) \geq 1 \text{ pour } i = 1, \dots, M$$

avec  $y_i = 1$  si  $\mathbf{x}_i \in$  classe 1,  $y_i = -1$  si  $\mathbf{x}_i \in$  classe 2,  $\mathbf{w}$  vecteur des poids qui est perpendiculaire à l'hyperplan,  $\omega_0$  le seuil et  $M$  le nombre d'images exemples. La méthode standard pour résoudre ce problème est d'utiliser la méthode de Lagrange. Si les images exemples ne sont pas linéairement séparables, on peut relaxer la contrainte et introduire un terme de coût supplémentaire dans la fonction coût. La fonction coût devient

$$F(\mathbf{w}) = \mathbf{w}^T \mathbf{w} / 2 + C \sum_{i=1}^M \xi_i$$

et les nouvelles contraintes

$$C_1 : \mathbf{w}^T \cdot \mathbf{x}_i + \omega_0 \geq 1 - \xi_i \text{ pour } y_i = 1$$

$$C_2 : \mathbf{w}^T \cdot \mathbf{x}_i + \omega_0 \leq -1 + \xi_i \text{ pour } y_i = -1$$

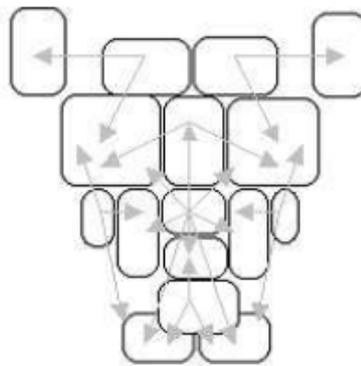
$$C_3 : \xi_i \geq 0$$

avec  $C$  un paramètre de régulation et  $\xi_i$  une variable d'erreur. Si les images exemples ne sont pas linéairement séparables, une autre méthode consiste à projeter les images exemples par des fonctions non-linéaires  $\Phi(\mathbf{x})$  dans un espace dans lequel ils sont

linéairement séparables et à appliquer l'algorithme SVM dans cet espace. Il est conseillé d'utiliser des fonctions non-linéaires  $\Phi(x)$  simples, car sinon il est probable de créer un « overfitting » des données.

## 6.4. Template matching

Les templates peuvent être définis soit "manuellement", soit paramétrés à l'aide de fonctions. L'idée est de calculer la corrélation entre l'image candidate et le template. Ces méthodes rencontrent encore quelques problèmes de robustesse liés aux variations de lumière, d'échelle, etc. Sinha [48, 49] utilise un ensemble d'invariants décrivant le modèle du visage. Afin de déterminer les invariants aux changements de luminosité permettant de caractériser les différentes parties du visage (telles que les yeux, les joues, et le front); cet algorithme calcule ainsi les rapports de luminance entre les régions du visage et retient les directions de ces rapports. La figure 15 montre un modèle prédéfini correspondant à 23 relations. Chaque flèche représente une relation entre deux régions. Une relation est vérifiée si le rapport entre les deux régions qui lui correspond dépasse un seuil. Le visage est localisé si le nombre de relations essentielles et de confirmation dépasse lui aussi un seuil.



**FIG 15:** Modèle de visage composé de 16 régions (les rectangles) associées à 23 relations (flèches). [48]

## 6.5. Le modèle de Markov caché

Les modèles de Markov cachés (HMM) sont utilisés depuis plusieurs années pour la détection et la reconnaissance du visage [50]. Cependant, les HMM de base sont utilisés de manière plus efficace si les données sont modélisées dans une seule dimension, ce qui

n'est pas le cas pour les images. Donc différentes variantes ont été proposées mais celle des

“Embedded HMM” génère des résultats supérieurs [51]. Les Embedded HMM [52] sont caractérisés par l'utilisation d'un HMM de base, modélisant l'apparence du visage de haut en bas. Ensuite, chacun des états de ce modèle général contient un autre HMM 1D, dénommé embedded (c à d. incorporé). Ceux-ci modélisent cette fois l'apparence du visage de la gauche vers la droite. Reposant sur certains coefficients de la transformée en cosinus discrète (DCT) comme source d'observations, les Embedded HMM constituent un algorithme de reconnaissance très performant. Or, les temps d'exécution des phases d'apprentissage et de test sont relativement élevés, nuisant donc à son utilisation en temps réel sur d'immenses banques d'images.

### 6.6. L'utilisation des points caractéristiques

La reconnaissance de visage basée sur les points caractéristiques est un processus de reconnaissance qui utilise les points caractéristiques: les yeux et la bouche comme une entrée aux algorithmes de reconnaissance. Ces algorithmes peuvent se servir de l'information appropriée telle que l'endroit ou les statistiques locales. Dans [53] KAMEL M. et al. ont utilisé des points caractéristiques du visage pour former un modèle géométrique du visage. Pour ce faire, ils ont extrait les 12 points suivants (figure 16) :

- Les quatre coins des yeux : A, B, C, et D.
- Le point S centre de gravité de A, B, C et D (milieu de la tête).
- Le point G à la base du nez.
- Les points extrême droit et extrême gauche du nez : E et F.
- Le point H milieu du contour supérieur de la lèvre supérieure.
- Les deux coins de la bouche I et J.
- Le M à l'intersection des droites (SG) et (IJ) représentant le milieu de la bouche.

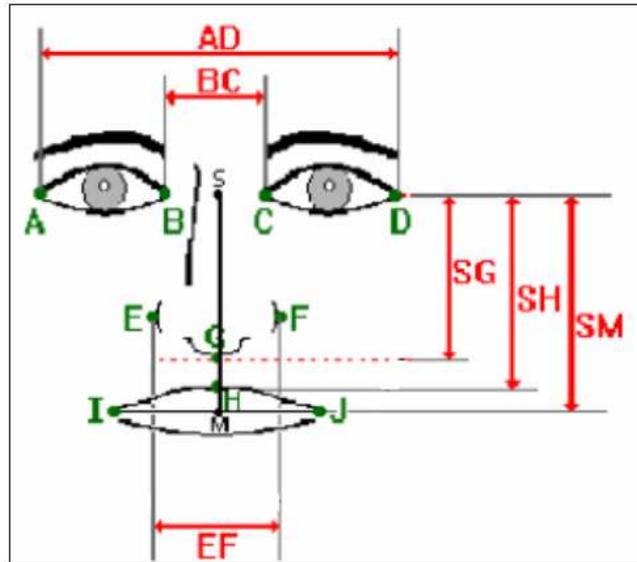


FIG 16: Modèle géométrique de visage

A partir de ces 12 points, ils calculent un vecteur caractéristique pour chaque visage, puis ils utilisent la distance euclidienne avec la méthode des (k-ppv) pour calculer la distance entre le vecteur caractéristique du visage à identifier et chaque vecteur. A partir des distances obtenues, et d'un seuil de rejet, ils reconnaissent le visage ou le considèrent comme inconnu.

## 6.7. Appariement élastique des graphes

Lades et al. [54] ont proposé un graphe topologique élastique connu sous le nom de « Elastic Graph Matching ». Les auteurs utilisent pour cela une grille régulière, placée sur les images de visages. Les caractéristiques extraites sont des coefficients de Gabor. Les nœuds de la grille sont liés élastiquement. L'élasticité permet d'accommoder des distorsions d'objet et des changements de point de vue. Ce système peut s'adapter aux changements modérés de taille et d'orientation du visage. La figure 17 (a) montre un exemple d'un graphe de visage. Depuis son application, l'EGM pour la reconnaissance de visage est devenu un champ de recherche très actif. Dans [55], l'auteur a montré que l'EGM surpasse, en termes de performances, l'ACP et des réseaux neurones appliqués à l'identification du visage. Dans [56], l'approche du graphe a été améliorée en introduisant une structure de pile pour chaque nœud. En d'autres termes, chaque nœud contient des réponses des filtres avec différents états du nœud. Cette approche est appelée "ElasticBench graph matching" (EBGM). La figure 18 (b) montre un exemple d'un EBGM

d'une image de visage. Par exemple, dans le cas de nœuds identifiant les coins de la bouche, les réponses des filtres de Gabor seront calculées sur la bouche ouverte, fermée, souriante..., et ceci pour tous les nœuds.

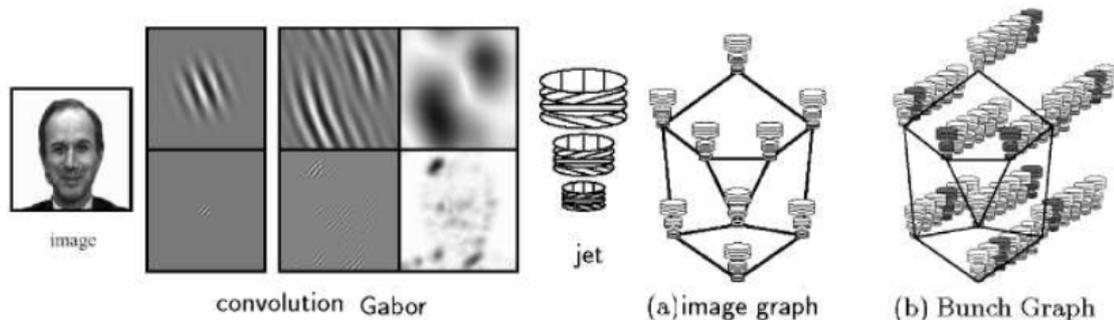


FIG 17: Structure des approches EGM et EBM

## 7. Comparaison des algorithmes

Les algorithmes de reconnaissance de visages précédemment décrits ne sont pas tous équivalents. En effet, outre le fait qu'ils se basent sur des approches d'analyse différentes, leurs performances de reconnaissance et leur robustesse face aux variations sont différentes. Le tableau 2avantages et les inconvénients de chaque algorithme.

Algorithme	Les avantages	Les inconvénients
<b>Analyse en composantes principales</b>	<ul style="list-style-type: none"> <li>- Flexibilité</li> <li>- Puissance</li> <li>- Puissance des résultats</li> <li>- Simplicité mathématique</li> </ul>	<ul style="list-style-type: none"> <li>- Coût de calcul élevé</li> <li>- Détermination du nombre de visages propres</li> </ul>
<b>Les réseaux de neurones</b>	<ul style="list-style-type: none"> <li>- Résistance au manque de fiabilité des données.</li> <li>- Dépendent peu de connaissance préalable</li> <li>- La capacité de représenter n'importe quelle dépendance fonctionnelle</li> </ul>	<ul style="list-style-type: none"> <li>- choix des données d'apprentissage, choix de type de réseau et de fonction</li> <li>- Apprentissage long</li> <li>- Sur-apprentissage</li> </ul>

<b>Les machines à vecteur de supports</b>	<ul style="list-style-type: none"> <li>- Flexibilité des noyaux.</li> <li>- L'apprentissage facile à réaliser</li> </ul>	<ul style="list-style-type: none"> <li>- un nombre élevé d'exemples peut être requis pour obtenir une bonne classification</li> </ul>
<b>Template matching</b>	<ul style="list-style-type: none"> <li>- Robuste pour localiser les régions</li> </ul>	<ul style="list-style-type: none"> <li>-Complexité de calcul</li> <li>-Espace mémoire important</li> </ul>
<b>le modèle de Markov caché</b>	<ul style="list-style-type: none"> <li>- Algorithmes d'apprentissage très robuste</li> </ul>	<ul style="list-style-type: none"> <li>- Temps d'apprentissage long</li> </ul>
<b>l'utilisation des points caractéristiques</b>	<ul style="list-style-type: none"> <li>- Requiert moins d'espace mémoire</li> <li>- Plus rapide et simple à implémenter</li> </ul>	<ul style="list-style-type: none"> <li>-Dépendent étroitement de l'exactitude d'extraction des caractéristiques</li> <li>-Prise en charge de la variation de la pose, de l'éclairage</li> </ul>

**Tableau 2:** Comparaison entre les algorithmes de reconnaissance de visage

## 8. Conclusion

La reconnaissance de visage est une technologie biométrique qui est trop utilisé dans les applications de contrôle de frontière, dans la sécurité des établissements et des zones urbaines et dans l'identification des conducteurs. C'est une technique commune, populaire, simple qui offre beaucoup d'avantages. Une multitude d'algorithmes globales, locales et hybrides sont utilisés et développés dans les systèmes commerciaux et dans les projets de recherche. Un système automatique de reconnaissance de visages est basée spécialement sur deux modules: l'extraction de caractéristiques et l'apprentissage. Le module d'extraction de caractéristiques à une influence notable sur la performance globale d'un système de reconnaissance de visage, cela est dû à la complexité des données dont on dispose d'une part et à la diversité des techniques utilisées dans ce module d'autre part. les méthodes utilisées pour les deux module sont décrites dans les deux chapitre suivants.

## Chapitre III

# SVM pour la reconnaissance de visages

### 1. Introduction

Depuis quelques années, de nouvelles méthodes d'apprentissage se développent sur la base de la Théorie de l'Apprentissage Statistique (Statistical Learning Theory) de Vapnik et Chervonenkis [57]. L'une de ces méthodes, appelée Machine à Vecteur de Support ou séparateur à vaste marge (en anglais Support Vector Machine ou SVM), permet de réaliser des estimations en classification, c'est-à-dire décider à quelle classe appartient un échantillon [58], ou en régression, c'est-à-dire prédire la valeur numérique d'une variable [59]. La classification par SVM est une méthode très puissante qui a démontré de très bons résultats dans plusieurs domaines notamment dans la reconnaissance des visages et des caractères manuscrits. Dans ce chapitre nous avons présenté un aperçu sur les SVM avec une relecture des travaux ultérieurs qui montre que l'application de cette méthode dans la reconnaissance des visages a permis un taux très intéressant de classification.

### 2. Aperçu sur les SVM

L'origine des machines à vecteurs de support remonte à 1975 lorsque Vapnik et Chervonenkis proposèrent le principe du risque structurel et la dimension VC pour caractériser la capacité d'une machine d'apprentissage. A cette époque, ce principe n'a pas trouvé place et il n'existait pas encore un modèle de classification solidement appréhendé

pour être utilisable. Il a fallu attendre jusqu'à l'an 1982 pour que Vapnik propose un premier classificateur basé sur la minimisation du risque structurel nommé SVM. Ce modèle était toutefois linéaire et l'on ne connaissait pas encore le moyen d'induire des frontières de décision non linéaire. En 1992, Boser et al proposent d'introduire des noyaux non linéaires pour étendre le SVM au cas non linéaire. En 1995, Cortes et al proposent une version régularisée du SVM qui tolère des erreurs d'apprentissage tout en les pénalisant. Depuis, les SVM n'ont cessé de susciter l'intérêt de plusieurs communautés de chercheurs de différents domaines d'expertise [60].

## 2.1. Théorie de base des SVM

Pour un problème de classification à deux classes [57], l'objectif est de séparer les deux classe par une fonction qui est induite à partir des exemples disponibles. Considérons l'exemple à la figure 18(a), où il ya beaucoup de séparateur possible qui peuvent séparer les données, mais il n'y a qu'un seul (figure 18(b)) qui maximise la marge (la distance entre l'hyperplan et les point les plus proches des données de chaque classe). Ce séparateur est appelé un hyperplan optimal.

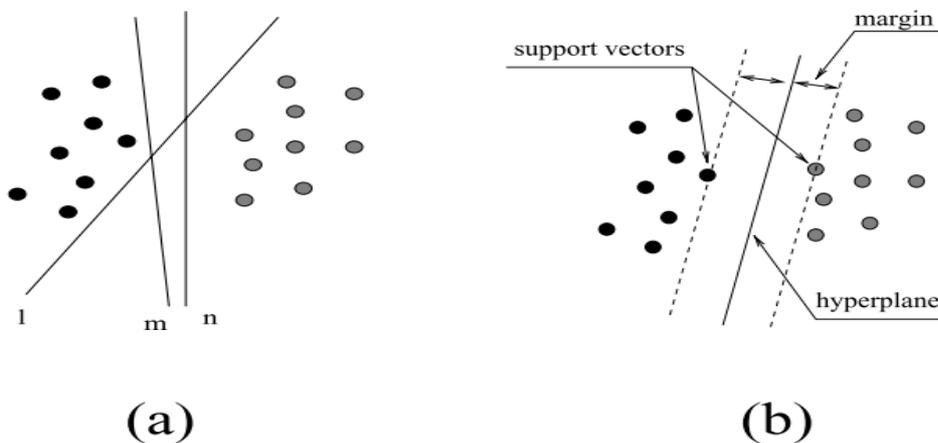


FIG 18 : Classification entre deux classes

Etant donné un ensemble des vecteurs d'apprentissage appartient à deux classes distinctes  $(x_i, y_i) \ i=1, \dots, N$  où chaque  $x_i \in \mathbb{R}^n$  et  $y_i \in \{-1, 1\}$ ,  $y_i$  définissant la classe de  $x_i$ , avec un hyperplan  $wx+b=0$ . L'ensemble des vecteurs est dite optimale séparée, si elle séparée sans erreur et la marge est maximale. Un hyperplan canonique a la contrainte pour  $w$  et  $b$ :  $\min_{x_i} (wx_i + b) = 1$ . Un hyperplan en forme canonique doit satisfaire la contrainte suivante:

$$y_i (w x_i + b) \geq 1, \quad i=1, \dots, l \quad (1)$$

la distance d'un point  $x$  de l'hyperplan est:

$$d(w, b, x) = \frac{lw x + bl}{llwll} \quad (2)$$

Pour cet hyperplan la marge entre les deux classes vaut  $\frac{2}{llwll}$ .

L'hyperplan optimal peut être obtenu par la solution du problème d'optimisation quadratique suivant :

$$\left\{ \begin{array}{l} \min \frac{1}{2} llwll^2 \\ \forall i \quad y_i (w x_i + b) \geq 1 \end{array} \right. \quad (3)$$

L'introduction des multiplicateurs de Lagrange  $\alpha_i > 0, i=1 \dots l$  donne le lagrangien primaire qui s'écrit :

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^l \alpha_i \{ y_i [(w \cdot x_i) + b] - 1 \} \quad (4)$$

Le lagrangien doit être minimal par rapport à  $w, b$  et maximal par rapport à  $\alpha$ , Le point optimal est un point qui vérifie :

$$\begin{array}{l} \frac{\partial L(w, b, \alpha)}{\partial w} = 0 \\ \frac{\partial L(w, b, \alpha)}{\partial b} = 0 \end{array} \quad \Longrightarrow \quad \begin{array}{l} W = \sum_{i=1}^l \alpha_i y_i x_i \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{array} \quad (5)$$

En substituant  $w$  par son expression (5) dans (4) on peut écrire le la problème quadratique dual d'optimisation sous la forme :

$$\left\{ \begin{array}{l} \text{Max} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i, x_j) \\ \forall i \quad \alpha_i \geq 0 \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{array} \right. \quad (6)$$

Par résolution du problème dual d'optimisation (6) on obtient les coefficients  $\alpha_i$  nécessaires pour l'expression du vecteur  $w$  (5) et on peut donc, établir la fonction de décision suivante :

$$f(x) = \sum_{i=1}^l \alpha_i^* y_i (x, x_i) + b^* \quad (7)$$

Les points  $x_i$  avec  $\alpha_i > 0$  sont appelés vecteurs de supports (VS). Dans le cas où l'hyperplan linéaire dans l'espace d'entrée ne sépare pas bien les données des deux classes, il est possible de créer un hyperplan qui offre une séparation linéaire dans un espace de redescription de grande dimension, éventuellement infinie, il devient donc possible d'envisager d'utiliser la méthode des SVM. Notons  $\Phi$  une transformation non linéaire de l'espace d'entrée  $X$  en un espace de redescription  $\Phi(x)$  :

$$X = (x_1, \dots, x_d)^T \text{ à } \Phi(x) = (\Phi(x_1), \dots, \Phi(x_d))^T \quad (8)$$

Généralement, le vecteur image  $\Phi(x)$  est de dimension supérieure à  $d$ , la dimension de l'espace d'origine. Les données sont projetées via la fonction  $\Phi$  telle que :

$\Phi(x_i) \cdot \Phi(x) = k(x_i, x)$ , où  $k$  est appelée fonction noyau ou (kernel). La figure montre que les données non linéairement séparables dans l'espace d'entrée  $X$  sont à présent séparables dans l'espace augmenté  $\Phi(x)$ .

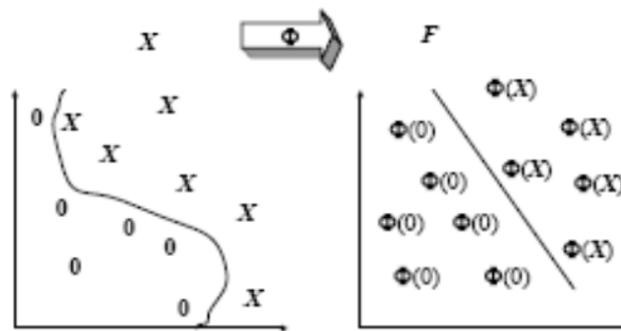


FIG 19 : Illustration de l'effet de changement d'espace par une fonction

Parmi les noyaux couramment utilisés pour la classification on peut citer:

- Noyau polynomiale  $K(x,y) = (x \cdot y + c)$  (9)
- Noyau gaussien (RBF)  $K(x,y) = \exp(-\|x - y\|^2 / \sigma^2)$  (10)
- Noyau sigmoïde  $K(x,y) = \tanh((a(x,y) - b))$  (11)

Le problème d'optimisation (6) se transcrit dans ce cas par :

$$\begin{cases} \text{Max}_{\alpha_i} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \forall i \quad 0 \leq \alpha_i \leq C \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \quad (12)$$

La fonction de décision dans le nouvel espace devient :

$$f(x) = \sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^* \quad (13)$$

Où les coefficients  $\alpha_i^*$  et  $b^*$  sont obtenus comme précédemment par résolution de l'équation (12).

## 2.2. Classification des données multi-classe

A l'origine les SVM ont été conçus essentiellement pour qui ne traite que des données appartenant à deux classes. Cependant plusieurs approches permettant d'étendre cet algorithme aux cas à N classes ont été proposées [61]:

### ➤ Approche « Un Contre Tous »

L'approche « Un Contre Tous » [62] est la plus simple et la plus ancienne des méthodes de décomposition. L'idée de cette stratégie est de construire autant de classifieurs que de classes. Le  $k^{\text{ième}}$  classifieur est destiné à distinguer la classe d'indice  $k$  de toutes les autres. Durant l'apprentissage, tous les exemples appartenant à la classe considérée sont étiquetés positivement (+1) et tous les exemples n'appartenant pas à la classe sont étiquetés négativement (-1). A la fin de l'apprentissage, nous disposons de  $K$  modèles correspondant aux hyper-plans  $(w_i, b_i)$  tel que  $i=1, \dots, K$ .

### ➤ Approche « Un Contre Un »

L'approche « Un Contre Un » [63] est un cas spécial des méthodes de décomposition proposées par Dietterich et al pour des problèmes à plusieurs classes. Cette approche consiste à utiliser un classifieur par couple de catégories. Le classifieur indicé par le couple  $(k, l)$  avec  $1 \leq k < l \leq K$  est destiné à distinguer la catégorie d'indice  $k$  de celle d'indice  $l$ . Pour la classification, un vecteur d'entrée  $X$  est présenté à  $\frac{1}{2} K(K-1)$  SVM.

## 3. Techniques de reconnaissance de visages à base de SVM

### 3.1. NEAR SET et SVM pour la reconnaissance de visage

Le système de reconnaissance de visage de Shivani et al [64] est composé de deux grandes phases: l'extraction de caractéristiques et la reconnaissance. Pour la première phase Shivani et al [64] ont développé une procédure automatisée pour extraire les traits de visage. Un prétraitement des visages est effectué pour obtenir les régions d'intérêt des visages et pour extraire les caractéristiques du visage, ils ont utilisé les traits de visage à savoir: la longueur du nez, la largeur du nez et la

distance entre les globes oculaires. Shivani et al [64] ont utilisé le concept de Near set pour choisir la fonction entre ces trois caractéristiques qui permet d'obtenir le meilleur taux de reconnaissance de visage. Pour la deuxième phase ils ont utilisé les SVM comme classificateur de visage. L'expérience a été réalisée sur la base de visage ORL, en utilisant le noyau RBF avec le paramètre  $\gamma=32$  et  $C=100$ , ils ont réduit le taux d'erreur à 0,2%.

### 3.2. Caractéristiques de Gabor et SVM pour la reconnaissance de visage

Linlin SHEN [65] propose une méthode de reconnaissance de visage basée sur les SVM qui utilise les caractéristiques de Gabor optimisées. Le système commence par l'extraction de caractéristiques en utilisant la méthode de Gabor. Les caractéristiques Gabor extraites et leur classes associées pour tous les exemples d'apprentissage sont introduites dans un algorithme de boosting pour éliminer les caractéristiques non discriminatives qui ne sont pas significatives pour la classification. Pour l'évaluation du système Linlin SHEN [65] a utilisé la base de données FERET (voir annexe A). Deux images pour chaque personne sont utilisées pour l'apprentissage et une pour le test. Le taux de reconnaissance obtenu est de 92% avec le noyau linéaire et en utilisant 120 caractéristiques de Gabor.

### 3.3. PCA, LDA et SVM pour la reconnaissance de visage

Jan Mazanec et al [66] présentent les résultats des différents algorithmes, à savoir: l'ACP (Analyse en Composantes Principales), LDA (Analyse discriminante linéaire) et SVM (Support Vector Machines). PCA vise à maximiser la séparation entre les classes de données, tandis que LDA essaie de maximiser la séparation entre les classes de données et de minimiser la séparation dans les classes. LDA utilise la matrice de covariance  $V$  obtenue à partir de PCA comme donnée d'entrée. L'avantage est la coupe des vecteurs propres de la matrice  $V$  qui ne sont pas importants pour la reconnaissance de visage. Lors des tests, Jan Mazanec et al [66] ont utilisé SVM avec le noyau RBF (radial base function). Les tests ont été effectués sur un groupe de 155 niveaux de gris sélectionnés à partir de bases de données d'images FERET. Ils ont utilisé trois différents ensembles d'images pour

l'apprentissage, soit: 2, 3 et 4 images par sujet dans l'ensemble d'apprentissage, tandis que le reste des images ont été utilisées pour les tests. Ils ont examiné cinq configurations différentes de la reconnaissance de visage (*SVM*, *PCA*, *LDA*, *PCA+SVM*, *LDA+SVM*). Les résultats montrent que la combinaison *LDA+SVM* donne les meilleurs résultats avec un taux de reconnaissance proche de 100%.

### 3.4. Transformée d'ondelettes et SVM pour la reconnaissance de visage

Bing et al [67] ont proposé un nouveau système de reconnaissance de visage qui utilise la transformée d'ondelette combiné avec la méthode SVM et la méthode de clustering. Ils ont utilisé la sous-bande LL de deux niveaux de la décomposition en ondelettes pour une entrée de taille  $28 \times 23$  de SVM. Pour réduire le calcul ils ont proposé un clustering hiérarchique. Les Tests ont été effectués sur la base de données ORL. L'ensemble d'apprentissage a été mis en place par une sélection aléatoire de 3 et 5 échantillons par personne à partir de la base de données entière et l'ensemble des tests concernait les images restantes. Les taux de reconnaissance obtenus sont de 94,8% et 98,3%. Les résultats expérimentaux montrent que cette approche a augmenté la précision de reconnaissance de visage et est robuste à l'illumination et l'expression.

### 3.5. KPCA et SVM pour la reconnaissance de visage

Ivanna et al [68] présentent un système de reconnaissance de visage basé sur la combinaison de deux méthodes KPCA et SVM. La méthode KPCA est un développement de la méthode PCA, elle est utilisée pour traiter la non-linéarité du problème de reconnaissance de visage. En utilisant une fonction noyau non linéaire, une réduction dimensionnelle est réalisée. Les SVM sont ensuite appliqués sur les caractéristiques extraites pour classer de nouvelles images de visage. Les expérimentations sont faites sur la base de données VISIO. La méthode KPCA+SVM est comparée avec d'autres méthodes telles que: la méthode NN (Nearest Neighbor), la méthode SVM (sans utilisation de KPCA pour l'extraction de caractéristiques), et la combinaison SVM+NN. Les résultats montrent que la méthode KPCA+SVM est plus performante que les autres méthodes avec un taux de reconnaissance de 99,05% en utilisant  $C=500000$ .

### 3.6. Bayesian et SVM pour la reconnaissance de visage

Contrairement aux méthodes traditionnelles de reconnaissance de visage basées sur les SVM qui utilisent un grand nombre de SVM pour l'apprentissage, Zhifeng Li et Tang Xiaoou[69] ont développé une nouvelle approche combinant la méthode d'analyse bayésienne avec les SVM. La méthode bayésienne convertit efficacement le problème de reconnaissance de visage multi classe en un problème de classification à deux classes, ce qui convient pour l'utilisation directe des SVM. Cette méthode n'a besoin que d'un seul SVM dans l'apprentissage pour classifier la différence entre la variation intra-classe et la variation extra-classe. la nouvelle méthode a aussi un inconvénient inhérent, elle essaie de séparer deux sous-espaces complexes par un seul hyperplan. Pour améliorer les performances de reconnaissance Zhifeng Li et Tang Xiaoou[69] ont encore développé trois méthodes s'inspirant de la méthode bayésienne basée sur les SVM: la méthode un contre tous, la méthode HAC, et la méthode de clustering adaptatif. Ils ont effectué des expériences sur les deux bases de données FERET et XM2VTS. La meilleure précision 98,6% est obtenue en utilisant la base de données FERET avec 2 images de 495 personnes pour l'apprentissage et les 700 autres personnes pour le test.

### 3.7. Ondelette, ACP et SVM pour la reconnaissance de visage

Pour augmenter la précision des systèmes de reconnaissance de visage Mazloom et al [70] ont présenté une approche hybride en utilisant une combinaison d'ondelette, PCA, et SVM. Pour le prétraitement et l'extraction de caractéristiques ils ont appliqué la transformée d'ondelette et ACP. Durant la phase de classification les SVM avec la stratégie de l'arbre binaire sont appliqués pour résoudre le problème de reconnaissance de visage multi classe, avec une décision robuste en présence de grande variation de visages. Deux bases de données sont utilisées pour évaluer la méthode proposée, l'une est la base ORL et l'autre est une base de données plus large de 900 images de 90 personnes. Dans leurs expériences sur la base ORL, Mazloom et al [70] ont sélectionné 200 images (5 images par personne) pour l'ensemble d'apprentissage et les 200 images restantes sont utilisées comme ensemble de tests. L'approche a diminué le taux d'erreur à 2,57%. Dans leur deuxième expérience sur la base de données de 900 images, ils ont sélectionné 450

images (5 images par personne sont choisies au hasard) pour l'ensemble d'apprentissage et 450 images restantes pour le test, le taux d'erreur minimum obtenu est de 7,56% qui est meilleur que les autres méthodes.

### 3.8. Visages propres et SVM pour la reconnaissance de visage

Ergun et al [71] ont introduit une technique hybride pour la reconnaissance de visage composée de deux parties principales: l'extraction de caractéristiques et la classification. Dans la première partie Ergun et al [71] ont utilisé la méthode de visages propres qui est basée sur l'analyse en composante principale. Dans la deuxième partie, après la génération des vecteurs caractéristiques les SVM sont utilisés. Ils ont examiné la précision de la classification en fonction de trois types de noyaux SVM, pour le test ils ont utilisé la base de données ORL (240 images pour l'apprentissage et 160 images pour le test (6 images pour chaque personne)). Pour une classification simple ils ont utilisé uniquement les 40 premiers éléments de vecteur caractéristique. Le meilleur taux de reconnaissance obtenu est de 91,4% en utilisant le noyau RBF avec la valeur  $\sqrt{5}$  de sigma.

### 3.9. GMM et SVM pour la reconnaissance de visage

Hervé et al [72] ont présenté un nouvel algorithme de reconnaissance de visage basé sur les GMM et les SVM. Au début une étape de détection des visages est effectuée en utilisant l'algorithme *visual front-end*. Une fois que le visage est détecté, il est normalisé à la taille de visage propre. La décomposition du visage détecté sur les visages propres est calculée et utilisée comme caractéristiques du visage. Durant l'apprentissage, un modèle de mélange Gaussien client (GMM) est adapté en utilisant les caractéristiques extraites de chaque image de la Vidéo, ensuite une machine à vecteurs de supports (SVM) est utilisée pour trouver une frontière de décision entre le GMM client et le GMM pseudo-imposteurs. Lors du test, un GMM est adapté à la vidéo de test et une décision est prise à l'aide du SVM client appris antérieurement. Dans leurs expériences Hervé et al [71] ont utilisé la base de données BANCA audiovisuel comportant 52 locuteurs divisée en deux groupes G1 et G2 de 26 locuteurs chacun (13 femmes et 13 hommes). Cette division en deux groupes différents permet d'utiliser G2 comme modèle

d'apprentissage et G1 lors du test (et réciproquement). Cet algorithme apporte une égalité de 3,5% taux d'erreur (EER).

#### 4. Conclusion

Dans ce chapitre nous avons présenté de manière simple et complète le concept de système d'apprentissage introduit par Vladimir Vapnik, les Machines à Vecteur de Support. C'est une technique d'apprentissage supervisée destinée à résoudre des problèmes de discrimination; c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression: c'est-à-dire prédire la valeur numérique d'une variable. Elle est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des ensembles de données. Nous avons exposé des différentes techniques de reconnaissance de visage à base des SVM tel que SVM+ACP, SVM+GMM, SVM+Ondelette,... Le grand avantage, par rapport aux autres techniques, est la capacité à généraliser la classification. Ainsi, cette méthode est adaptée aux applications présentant une grande variation intra-classes. Dans notre travail nous avons choisis de combiner la méthode SVM avec une nouvelle méthode appelée SIFT [73]. SIFT (Scale Invariant feature transform), que l'on peut traduire par « transformation de caractéristiques visuelles invariante à l'échelle », est un algorithme utilisé dans le domaine de la vision par ordinateur pour détecter et identifier les éléments similaires entre différentes images numériques. L'étape fondamentale de la méthode proposée par Lowe consiste à calculer ce que l'on appelle les « descripteurs SIFT » des images à étudier. Il s'agit d'informations numériques dérivées de l'analyse locale d'une image et qui caractérisent le contenu visuel de cette image de la façon la plus indépendante possible de l'échelle (« zoom » et résolution du capteur), du cadrage, de l'angle d'observation et de l'exposition (luminosité). Ainsi, deux photographies de la tour Eiffel auront toutes les chances d'avoir des descripteurs SIFT similaires, et ceci d'autant plus si les instants de prise de vue et les angles de vue sont proches. D'un autre côté, deux photographies de sujets très différents produiront selon toute vraisemblance des descripteurs SIFT très différents eux aussi (pouvoir discriminant). Cette robustesse, vérifiée dans la pratique, est une exigence fondamentale de la plupart des applications et explique en grande partie la popularité de la méthode SIFT.

## Chapitre IV

# Approche hybride pour la reconnaissance de visages à base de SIFT et SVM

### 1. Introduction et motivations

Durant la dernière décennie, la haute sécurité est devenue un problème majeur pour les installations gouvernementale (militaire, nucléaire, air port,..). L'identification de personne par la reconnaissance de visages donne un certain nombre de solutions à ce problème. Dans ce travail nous avons proposé un système de reconnaissance de visage basée sur une méthode nommée SIFT consiste à rechercher des points caractéristiques (appelés « features ») sur une image qui seront décrits chacun par des coordonnées (x et y), une orientation, une échelle, ainsi que 128 descripteurs. Le gros avantage de l'algorithme SIFT est qu'il est invariant à la translation, à la rotation et au changement d'échelle et robuste aux transformations affines en général (distorsions), aux changements d'angle de vue ( $\pm 30^\circ$ ), ainsi qu'aux changements de luminosité. Le problème posé par l'algorithme SIFT est qu'il nécessite un temps de calcul importants lors de mise en correspondance dans le cas où le nombre de points caractéristiques est élevé. Ce problème est traité par l'utilisation de la méthode SVM qui a l'avantage de la capacité à généraliser la classification. Ainsi, cette méthode est adaptée aux applications présentant une grande variation intra-classes. Nous avons pu en effet démontrer que l'utilisation de SIFT avec SVM et d'obtenir des résultats très informatifs.

Dans ce chapitre nous décrivons en détail les différentes étapes de notre approche basée sur les descripteurs SIFT les SVM.

## 2. L'architecture globale du Système

Le système est constitué des modules suivants:

Calcul des points d'intérêt et de leurs descripteurs SIFT, l'apprentissage, la mise en correspondance des points d'intérêts, et la classification par les SVM. La figure 20 illustre l'architecture globale du système, qui sera détaillée par la suite.

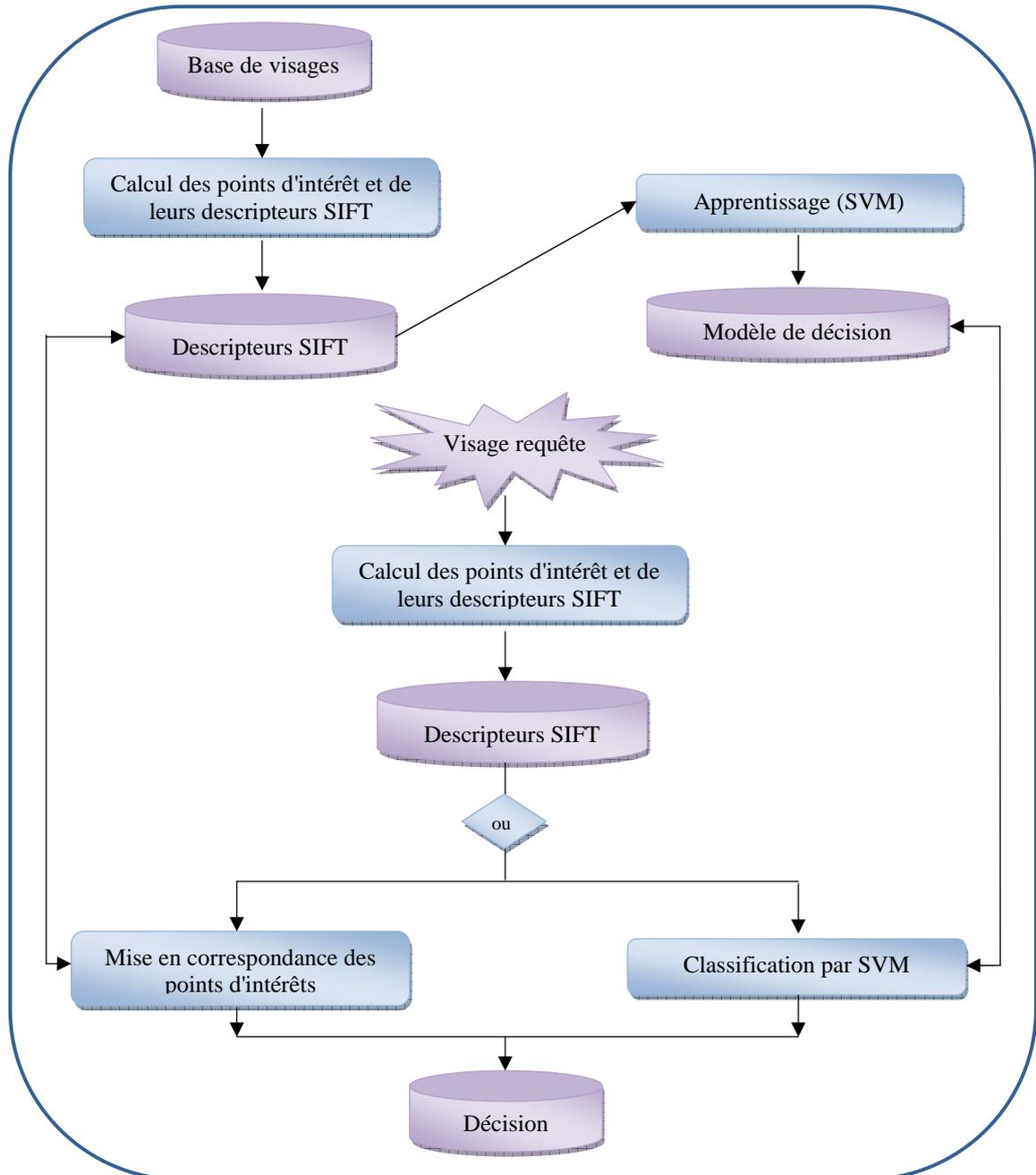


FIG20 : L'architecture globale du système

## 2.1. Calcul des points d'intérêt et de leurs descripteurs SIFT

SIFT [73] (Scale Invariant Feature Transform) est un descripteur de taille 128 et composé d'histogrammes d'orientations sur quatre fenêtres autour d'un point. Il est associé à l'orientation et l'échelle à laquelle il a été détecté. Les points sur lesquels ce descripteur est calculé; sont choisis en filtrant l'image par un filtre Gaussien à différentes échelles puis en extrayant les extrema en espace et en échelle. Ce descripteur est l'un des plus utilisés dans la littérature.

L'algorithme procède en quatre étapes principales : la détection des extremums (qui correspondent aux points d'intérêt) à un espace d'échelle crée par des filtre DoG sur une pyramide multi-échelle; la localisation et le filtrage des points d'intérêt; le calcul des histogrammes des directions des gradients locaux autour du point d'intérêt dans un voisinage qui varie en fonction de l'échelle du point, et le calcul d'une signature à partir des histogrammes en divisant la zone autour du point en quatre sous-zones.

### 2.1.1. Détection d'extremums d'espace-échelle (scale-space)

Pour la première étape de détection d'extremums d'espace-échelle, Lowe propose d'utiliser une structure d'échelles imbriquées afin d'analyser chaque point de l'image à diverses échelles, dans le but de ne retenir que l'échelle qui semble la plus pertinente. L'intérêt immédiat de ce raisonnement est définir un descripteur invariant par changement d'échelle. Reste ensuite à choisir la fonction d'échelle à proprement dite. D'après Lindeberg [74], le seul noyau espace-échelle possible est la Gaussienne  $G(x,y,\sigma)$ . On obtient par convolution l'image lissée à l'échelle  $\sigma$ :

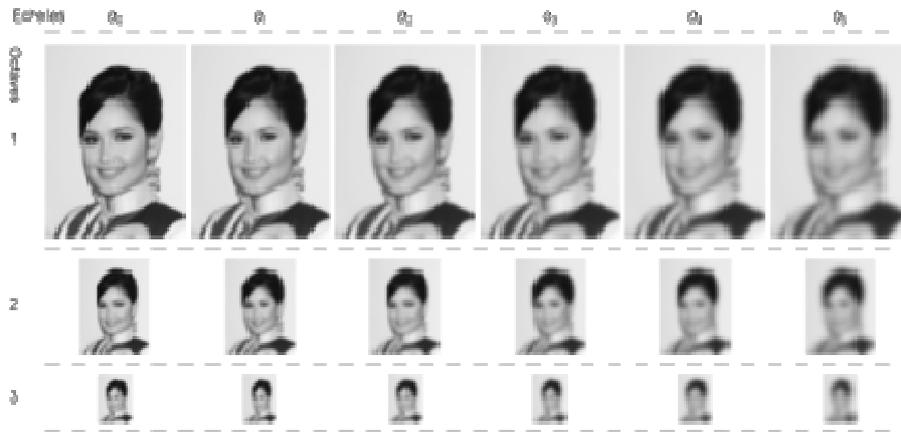
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

avec la fonction gaussienne à l'échelle  $\sigma$ :

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

Cette convolution a pour effet de lisser l'image originale  $I$  de telle sorte que les détails trop petits, c'est-à-dire de rayon inférieur à  $\sigma$ , sont estompés.

La Figure 21 nous montre les images Gaussiennes groupées par trois octaves. De gauche à droite, l'échelle augmente. De haut à bas, la taille d'image divisée par deux.



lients

Il faut utiliser cette image lissée pour chercher quels sont, à une échelle donnée, les points candidats au statut de point-clé. On comprend aisément que de tels points (qui sont représentatifs d'un objet de taille de l'ordre de grandeur de  $\sigma$ ) sont obligatoirement stables par rapport à  $\sigma$ , et donc que la dérivée de  $L(x, y, \sigma)$  par rapport à  $\sigma$  est nulle en  $(x, y)$  point-clé. Or seul  $G$  dépend de  $\sigma$  et on a :

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G$$

sachant que  $\sigma^2 \Delta G = \sigma^2 \nabla^2 G$ , le laplacien de  $G$  normalisé par  $\sigma^2$ , permet une véritable analyse invariante par changement d'échelle et l'analyse de ses extremums conduit à l'obtention de points-clés plus stables que ceux obtenus par d'autres critères (gradient, détecteur de coins de Harris ou Hessien). Partant de l'approximation suivante (taux de variation de  $G$  par rapport à la variable  $\sigma$ ) à partir de deux échelles  $\sigma$  et  $k\sigma > \sigma$  :

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

et en utilisant le Laplacien de  $G$  normé pour détecter les points candidats :

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1) \sigma^2 \nabla^2 G$$

on montre qu'il est équivalent d'étudier les extremums de la fonction  $D(x, y, \sigma)$  d'analyse espace-échelle définie par :

$$\begin{aligned}
 D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\
 &= L(x, y, k\sigma) - L(x, y, \sigma).
 \end{aligned}$$

Schématiquement, l'image analysée pour chaque échelle par la différence de gaussienne (DoG) est obtenue de la manière suivante (où une octave est une suite croissante d'échelle allant de  $\sigma$  à  $2\sigma$ ):

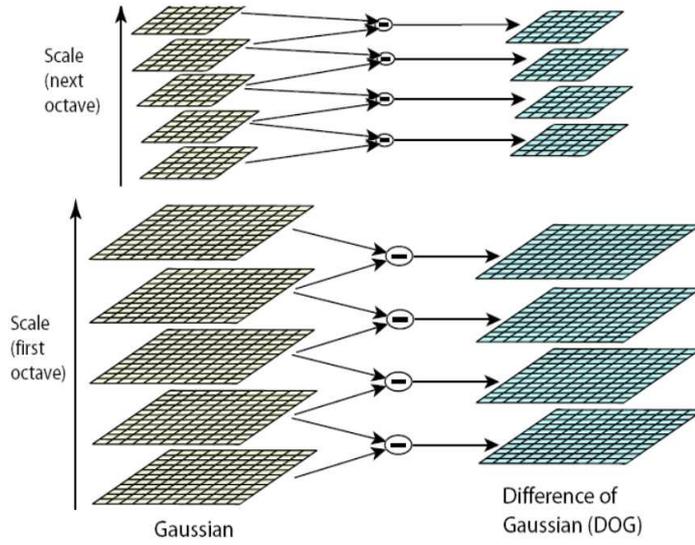


FIG22 : Construction de la fonction d'analyse multi-échelle de l'image

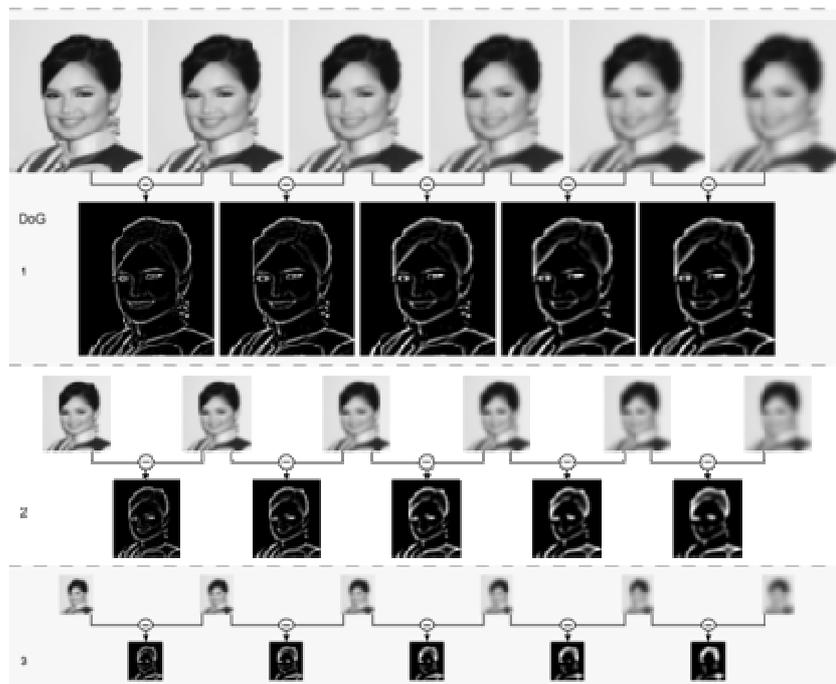


FIG23 : Construction de la pyramide de différences de gaussiens (DoG) à partir de la pyramide de gradients

Les extrema sont alors les pixels qui présentent une intensité maximum, ou minimum, par rapport à leurs huit plus proches voisins à la même échelle  $\sigma_j$  ainsi qu'à ceux dans l'espace-échelle (9 voisins dans l'échelle  $\sigma_{j-1}$  et 9 voisins dans l'échelle  $\sigma_{j+1}$ ) (voir FIG24).

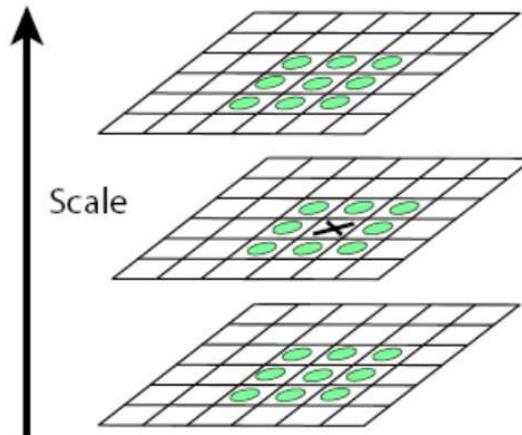


FIG24 : Comparaison aux 26 plus proches voisins

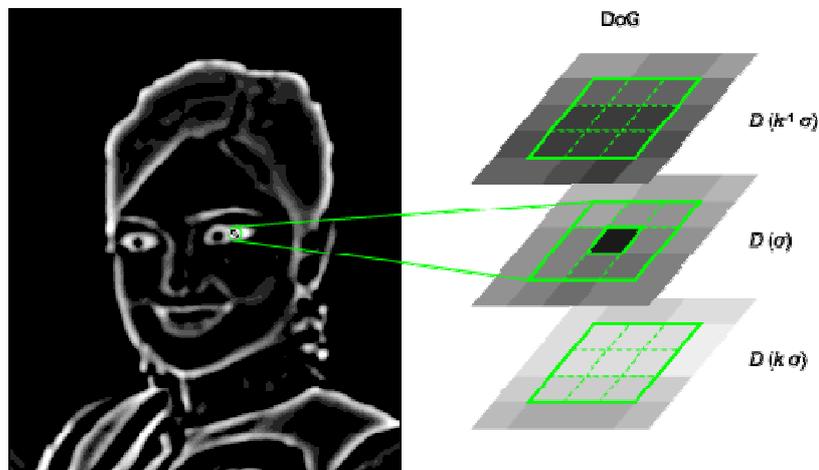


FIG25 : Exemple de détection d'extremums dans l'espace des échelles

### 2.1.2. Localisation des points d'intérêt

Visant à augmenter de façon significative la stabilité et la qualité de la mise en correspondance ultérieure, cette étape, qui est une amélioration de l'algorithme original, s'effectue dans l'espace des échelles à trois dimensions, où  $D(x, y, \sigma)$ , qui n'est connu que pour des valeurs discrètes de  $x$ ,  $y$  et  $\sigma$ , doit être interpolé.

Cette interpolation s'obtient par un développement de Taylor à l'ordre 2 de la fonction différence de gaussiennes  $D(x, y, \sigma)$ , en prenant comme origine les coordonnées du point-clé candidat. Ce développement s'écrit comme suit :

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

où  $D$  et ses dérivées sont évaluées au point-clé candidat et où  $\mathbf{x} = (x, y, \sigma)^T$  est un delta par rapport à ce point. Les dérivées sont estimées par différences finies à partir des points voisins connus de façon exacte. La position précise de l'extremum  $\hat{\mathbf{x}}$  est déterminée en résolvant l'équation annulant la dérivée de cette fonction par rapport à  $\mathbf{x}$  ; on trouve ainsi:

$$\hat{\mathbf{x}} = - \frac{\partial^2 D^{-1} \partial D}{\partial \mathbf{x}^2}$$

Un delta  $\hat{\mathbf{x}}$  supérieur à 0,5 dans l'une des trois dimensions signifie que le point considéré est plus proche d'un des voisins dans l'espace des échelles discret. Dans ce cas, le point-clé candidat est mis à jour et l'interpolation est réalisée à partir des nouvelles coordonnées. Sinon, le delta est ajouté au point candidat initial qui gagne ainsi en précision.

### 2.1.2.1. Test d'élimination des points-clés

Conserver tous les extremums retenus n'est pas en effet suffisant, car comme nous l'avons déjà évoqué, certains de ces points de sont pas pertinents (petits extremums locaux ou correspondant à des effets de bord).

#### 2.1.2.1.1. Les faibles réponses

Afin d'éliminer ces points qui seront sensibles au bruit, évaluons tout d'abord quelle est la valeur extrapolée de  $D$  en ce point. En utilisant la solution de la localisation de l'extremum dans l'expression de la fonction d'interpolation de  $D$  à partir de son développement de Taylor, on obtient:

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}}$$

Lowe a décidé de supprimer tous les points-clés ayant un module de  $D$  inférieur à 3% de la valeur maximum que peut prendre  $D$ .

### 2.1.2.1.2. Les bords

Les points situés sur les arêtes (ou contours) doivent être éliminés car la fonction DoG y prend des valeurs élevées, ce qui peut donner naissance à des extremums locaux instables, très sensibles au bruit : si l'image devait subir un changement numérique même imperceptible, de tels points-clés peuvent se retrouver déplacés ailleurs sur la même arête, ou même simplement disparaître.

Un point candidat à éliminer, si l'on considère les deux directions principales à sa position, est caractérisé par le fait que sa courbure principale le long du contour sur lequel il est positionné est très élevée par rapport à sa courbure dans la direction orthogonale. La courbure principale est représentée par les valeurs propres de la matrice hessienne  $\mathbf{H}$  :

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Les dérivées doivent être évaluées aux coordonnées du point d'intérêt  $(x,y,\sigma)$  dans l'espace des échelles. Les valeurs propres de  $\mathbf{H}$  sont proportionnelles aux courbures principales de  $D$ , dont seul le rapport  $r$  est intéressant. La trace de  $\mathbf{H}$  représente la somme de ces valeurs, le déterminant son produit. Par conséquent, en adoptant un seuil  $r_{th}$  sur le ratio des courbures ( $r_{th} = 10$  dans la méthode originale de Lowe), un point-clé candidat va être retenu, selon le critère adopté par Lowe, si:

$$R = \frac{\text{tr}(\mathbf{H})^2}{\det(\mathbf{H})} = \frac{(r + 1)^2}{r} < \frac{(r_{th} + 1)^2}{r_{th}}$$

La vérification de ce critère est rapide, ne nécessitant qu'une dizaine d'opérations flottantes seulement. Lorsque ce critère n'est pas vérifié, le point est considéré comme localisé le long d'une arête et il est par conséquent rejeté.

Cette étape est inspirée de la technique de détection de points d'intérêt par l'opérateur de Harris ; pour le seuillage, une matrice hessienne est utilisée au lieu de la matrice des moments d'ordre 2 (tenseur).

### 2.1.3. choix de l'orientation des descripteurs

Comme il l'a été déjà évoqué, l'invariance par rotation est primordiale car on doit pouvoir analyser un objet indépendamment de sa position et de son orientation dans l'image, et il doit être possible de comparer deux objets suivant une direction privilégiée sans quoi, à moins d'utiliser un descripteur invariant par rotation, deux objets identiques pourront être considéré comme différents.

Cette étape consiste à attribuer à chaque point-clé une ou plusieurs orientations déterminées localement sur l'image à partir de la direction des gradients dans un voisinage autour du point. Dans la mesure où les descripteurs sont calculés relativement à ces orientations, cette étape est essentielle pour garantir l'invariance de ceux-ci à la rotation : les mêmes descripteurs doivent pouvoir être obtenus à partir d'une même image, quelle qu'en soit l'orientation.

Pour un point-clé donné  $(x_0, y_0, \sigma_0)$ , le calcul s'effectue sur  $L(x, y, \sigma_0)$ , à savoir le gradient de la pyramide dont le paramètre est le plus proche du facteur d'échelle du point. De cette façon, le calcul est également invariant à l'échelle. À chaque position dans un voisinage du point-clé, on estime le gradient par différences finies symétriques, puis son amplitude (c.-à-d. sa norme)  $m(x, y)$ , et son orientation  $\theta(x, y)$ :

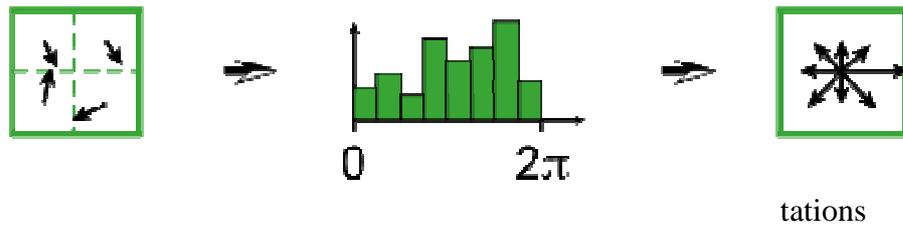
$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

$\forall (x, y)$  dans un voisinage de  $(x_0, y_0)$ .

On opère ainsi cette opération sur tous les points de la région définie par:

- ✓ son centre situé au plus proche voisin dans la grille espace du point-clé
- ✓ sa taille qui est liée à l'échelle la plus proche de celles utilisées pour l'étude de  $D(x, y, \sigma)$  Ensuite, on calcule l'histogramme des orientations des gradients en:
  - ✓ quantifiant les 360° en 36 cellules
  - ✓ ajoutant les contributions de chaque gradient par leur amplitude pondérée par la gaussienne de lissage de la région de variance  $1.5\sigma$

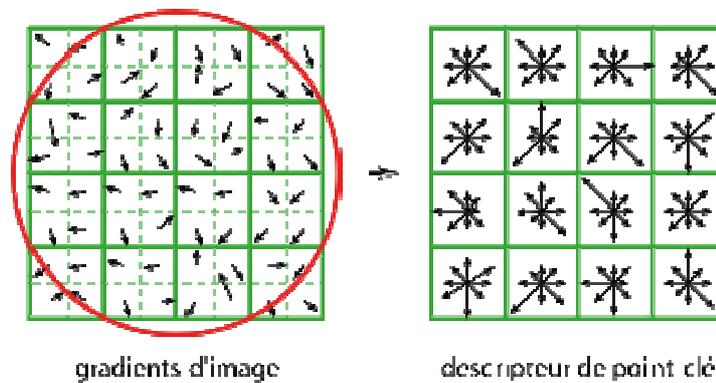


Les pics dans cet histogramme correspondent aux orientations dominantes. Toutes les orientations dominantes permettant d'atteindre au moins 80 % de la valeur maximale sont prises en considération, ce qui provoque si nécessaire la création de points-clés supplémentaires ne différant que par leur orientation principale.

À l'issue de cette étape, un point-clé est donc défini par quatre paramètres  $(x,y,\sigma,\theta)$ . Il est à noter qu'il est parfaitement possible qu'il y ait sur une même image plusieurs points-clés qui ne diffèrent que par un seul de ces quatre paramètres (le facteur d'échelle ou l'orientation, par exemple).

### 2.1.4. Calcul des descripteurs.

Une fois les points-clés, associés à des facteurs d'échelles et à des orientations, détectés et leur invariance aux changements d'échelles et aux rotations assurée, arrive l'étape de calcul des vecteurs descripteurs, traduisant numériquement chacun de ces points-clés. À cette occasion, des traitements supplémentaires vont permettre d'assurer un surcroît de pouvoir discriminant en rendant les descripteurs invariants à d'autres transformations telles que la luminosité, le changement de point de vue 3D, etc. Cette étape est réalisée sur l'image lissée avec le paramètre de facteur d'échelle le plus proche de celui du point-clé considéré.



Autour de ce point, on commence par modifier le système de coordonnées local pour garantir l'invariance à la rotation, en utilisant une rotation d'angle égal à l'orientation du point-clé, mais de sens opposé. On considère ensuite, toujours autour du point-clé, une région de  $16 \times 16$  pixels, subdivisée en  $4 \times 4$  zones de  $4 \times 4$  pixels chacune. Sur chaque zone est calculé un histogramme des orientations comportant 8 intervalles. En chaque point de la zone, l'orientation et la magnitude du gradient sont calculés comme précédemment. L'orientation détermine l'intervalle à incrémenter dans l'histogramme, ce qui se fait avec un double pondération – par l'amplitude et par une fenêtre gaussienne centrée sur le point clé, de paramètre égal à 1,5 fois le facteur d'échelle du point-clé.

Ensuite, les 16 histogrammes à 8 intervalles chacun sont concaténés et normalisés. Dans le but de diminuer la sensibilité du descripteur aux changements de luminosité, les valeurs inférieures à 0,2 sont remplacées par 0 et l'histogramme est de nouveau normalisé, pour finalement fournir le descripteur SIFT du point-clé, de dimension 128.

Cette dimension peut paraître bien élevée, mais la plupart des descripteurs de dimension inférieure proposés dans la littérature présentent de moins bonnes performances dans les tâches de mise en correspondance pour un gain en coût de calculs bien modéré, en particulier quand la technique *Best-Bin-First* (BBF) est utilisée pour trouver le plus proche voisin. Par ailleurs, des descripteurs de plus grande dimension permettraient probablement d'améliorer les résultats, mais les gains escomptés seraient dans les faits assez limités, alors qu'à l'inverse augmenterait sensiblement le risque de sensibilité à la distorsion ou à l'occlusion. Il a également été démontré que la précision de recherche de correspondance de points dépasse 50 % dans les cas de changement de point de vue supérieur à 50 degrés, ce qui permet d'affirmer que les descripteurs SIFT sont invariants aux transformations affines modérées. Le pouvoir discriminant des descripteurs SIFT a pour sa part été évalué sur différentes tailles de bases de données de points-clés ; il en ressort que la précision de mise en correspondance est très marginalement impactée par l'augmentation de la taille de la base de données, ce qui constitue une bonne confirmation du pouvoir discriminant des descripteurs SIFT.

## 2.2. Apprentissage (SVM)

Dans la partie apprentissage le système prend un ensemble d'exemples avec leurs classes connues à l'avance et construit un modèle de décision en utilisant la méthode SVM. Dans cette phase, on commence par l'extraction de leurs caractéristiques, puis leur enregistrement avec les classes correspondantes dans une base de caractéristiques.

La méthode SVM travaille sur la base des caractéristiques et essaye de trouver un modèle de décision. SVM exige que la base utilisée soit normalisée entre -1 et +1, pour cela, il est nécessaire de convertir les valeurs de tous les vecteurs dans l'intervalle  $[-1, +1]$ . Le principal avantage de cette normalisation est d'éviter les attributs de grandes valeurs en les dominant dans des petites intervalles. Un autre avantage est d'éviter les difficultés lors du calcul, parce que les valeurs du noyau dépendent généralement des produits scalaires des vecteurs de caractéristiques. On calcule alors le maximum et le minimum de chaque attribut de la table puis on applique la fonction suivante :

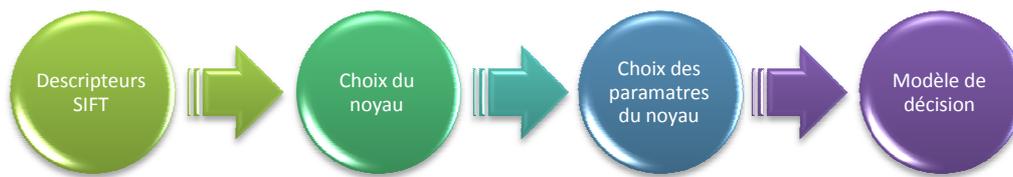
$$f: \mathbb{R} \rightarrow [-1, +1]$$

### Paramètres à varier

Durant l'apprentissage avec la méthode SVM, plusieurs noyaux peuvent être utilisés pour faire l'apprentissage et à chaque noyau correspondent différents paramètres à faire varier.

- Noyau linéaire : Aucun paramètre à varier.
- Noyau polynomial :  $K(x,y) = (x \cdot y + c)^n$ ,  $c \in \mathbb{R}$  et  $n \geq 2$  le paramètre important pour ce noyau est le degré du polynôme "**n**".
- Noyau RBF (Radial Basis Function):  $K(x,y) = \exp\left(-\|x - y\|^2 / \sigma^2\right)$ ,  $\sigma \in \mathbb{R}$ ,  $\sigma$  est le paramètre à faire varier pour ce noyau.

Le module SVM génère alors un modèle de décision pour chaque classe à partir de la base des caractéristiques normalisées (Figure 28).



**FIG28** : Apprentissage (SVM)

Pour chaque classe on prend la table normalisée et on met les valeurs de cette classe à +1 et toutes les autres à -1 puis on appelle le module SVM pour générer les paramètres de décision pour cette classe.

Les paramètres générés par SVM pour chaque classe sont :

- les  $\alpha_i$  différents de 0,
- $b_i$ ,
- les vecteurs des caractéristiques normalisés correspondants au  $\alpha_i$  différents de 0, c'est-à-dire les vecteurs supports.

Le modèle de décision global obtenu contient en plus des paramètres des classes :

- le noyau utilisé.
- les paramètres du noyau.
- Les libellés des classes.

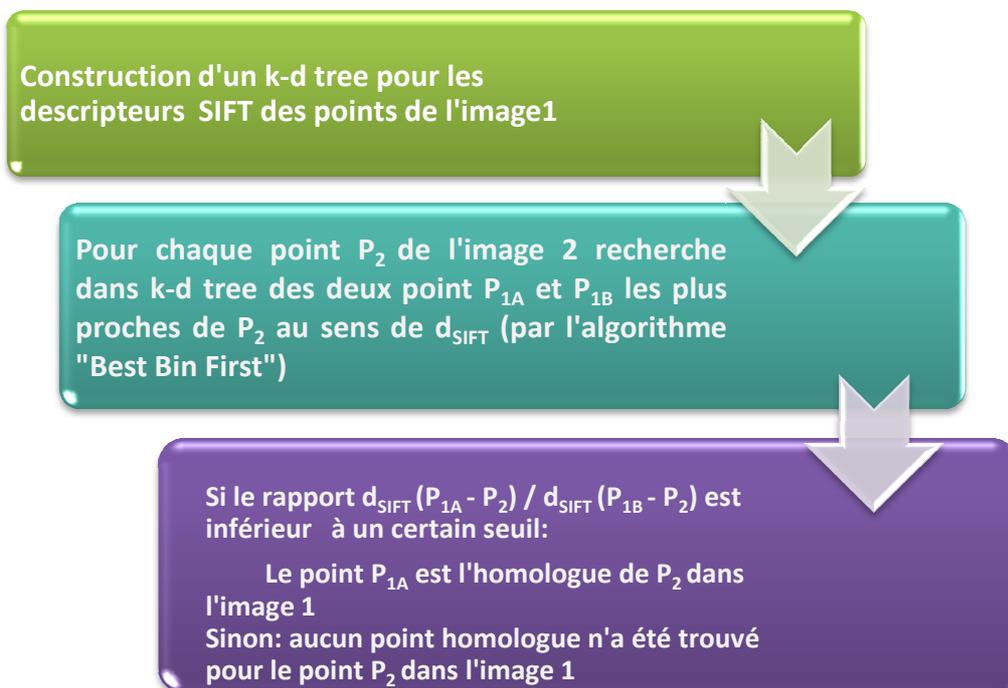
Le modèle obtenu peut être enregistré dans un fichier pour l'utiliser dans la phase de reconnaissance.

### 2.3. Mise en correspondance des points d'intérêts

L'indexation est l'opération de stockage des descripteurs SIFT des images de référence, d'une manière qui facilite l'identification des descripteurs correspondants de l'image question. Lowe utilise un arbre kd pour indexer les descripteurs puis une méthode de recherche dans cet arbre modifiée par rapport à l'approche classique, appelée *Best bin first*. Cette dernière est capable de trouver les plus proches voisins d'un descripteur question avec une bonne probabilité de façon très économe en temps de calcul. Cet algorithme se fonde sur deux astuces. Tout d'abord, le nombre de boîtes (feuilles de l'arbre kd) à explorer pour trouver le plus proche voisin d'un descripteur

question donné est limité à une valeur maximale fixée. Ensuite, les nœuds de l'arbre kd sont explorés dans l'ordre de leur distance au descripteur question, grâce à l'utilisation d'une file de priorité basée sur un tas binaire. Par distance d'un nœud à un descripteur, on entend distance euclidienne de la boîte englobant des feuilles sous-jacentes à ce descripteur. Dans la plupart des cas, ce procédé fournit la meilleure correspondance, et, dans les cas restants, un descripteur très proche de celle-ci.

De façon à déconsidérer les descripteurs faiblement discriminants (parce qu'ils sont trop souvent présents dans la base de référence, comme par exemple ceux qui sont extraits de motifs d'arrière-plan souvent répétés dans les images), Lowe recherche à la fois le plus proche voisin et le second plus proche voisin de chaque descripteur question. Lorsque le rapport des distances est supérieur à 0,8, la correspondance est éliminée, car considérée comme ambiguës (« bruit »). Par ce critère, Lowe parvient à éliminer 90 % des fausses correspondances en perdant moins de 5 % de correspondances correctes.



térêts

## 2.4. Classification par SVM

La reconnaissance est le module le plus important dans le système de reconnaissance de visage, elle utilise le modèle obtenu lors de la phase d'apprentissage pour décider à propos des nouvelles images. On prend une image de visage en entrée et on extrait ses caractéristiques avec la même méthode utilisé lors de la phase d'apprentissage, puis on applique pour chaque classe du modèle la fonction de décision qui donne une valeur réelle : si elle est positive le visage appartient à cette classe, si elle est négative le visage appartient à une autre classe.

Il peut arriver qu'un visage appartienne à plusieurs classes (plusieurs valeurs positives de la fonction), dans ce cas, le visage appartient à la classe qui maximise la fonction de décision. Cette méthode est appelée une contre reste et sa fonction de décision est :

$$k^* = \text{Arg max}_{1 \leq k \leq m} h_k(\mathbf{x})$$

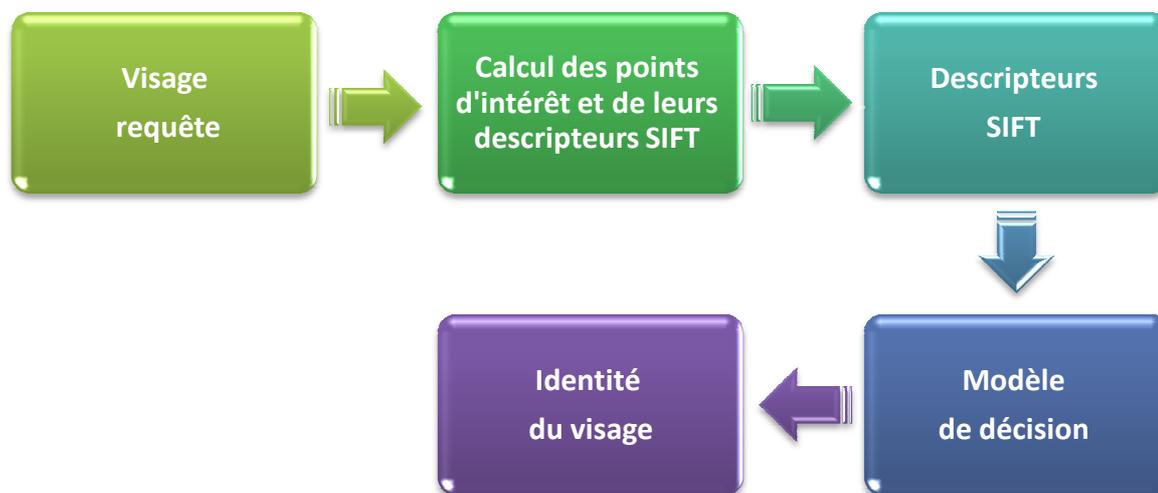


FIG30 : Classification par SVM

## 3. Conclusion

Nous avons proposé dans ce chapitre une méthode de reconnaissance de visage basée sur les descripteurs SIFT et l'algorithme SVM. L'avantage de cette méthode est qu'elle utilise comme caractéristiques des visages les descripteurs SIFT qui sont invariants à l'échelle, au cadrage, à l'angle d'observation et à l'exposition. Pour la classification nous

avons fait une comparaison entre la mise en correspondance des points d'intérêts et l'algorithme SVM qui s'est montré plus efficace dans les systèmes de reconnaissance de visages. Les détails de l'implémentation de l'approche proposée ainsi que les résultats obtenus et les bilans seront présentés dans le chapitre suivant.

# Chapitre V

## Mise en œuvre, bilans et résultats

### 1. Introduction

Dans ce chapitre, il sera question d'évaluer le système de reconnaissance de visages que nous avons réalisé et cela en fonction de la précision. Une large série d'expérimentations sera présentée et cela pour différents cas. Dans ces expérimentations, des testes ont été effectués sur des images contenant un ensemble de données avec une variabilité de la difficulté.

### 2. Réalisation du système

Le fichier image sur lequel nous allons effectuer des traitements est un fichier d'extension .pgm. En tête du fichier se trouve une zone dans laquelle sont enregistrées des informations concernant l'image. Puis, à la suite se trouvent tous les pixels formant l'image. Chaque pixel représentant le niveau de gris du pixel: 0 correspond au noir, 255 au blanc, les valeurs intermédiaires à des gris plus ou moins foncés.

Les fichiers i.pgm sur lesquels nous allons travailler ont la structure suivante:

P5	}	taille de l'image: 92 pixels de largeur x 112de hauteur Valeur max d'un pixel (blanc)
92 112		
255		

```

150 151 152 153 148 146 147 150 152 148 147 149 150 152 154 152 ..
150 151 152 153 148 146 147 150 152 148 147 149 150 152 154 154 ..
155 151 152 153 148 146 147 150 152 148 147 149 150 152 159 151 ..
155 151 152 153 148 146 147 150 152 148 147 149 150 152 160 153 ..
153 151 152 153 148 146 147 150 152 148 147 149 150 152 158 152 ..
153 151 152 153 148 146 147 150 152 148 147 149 150 152 155 154 ..
158 151 152 153 148 146 147 150 152 148 147 149 150 152 151 153 ..
158 151 152 153 148 146 147 150 152 148 147 149 150 152 157 151 ..
160 151 152 153 148 146 147 150 152 148 147 149 150 152 154 152 ..
160 151 152 153 148 146 147 150 152 148 147 149 150 152 152 155 ..
.....

```

Données binaires: niveau de gris de chaque pixel.

## 2.1. SIFT

Pour l'extraction des descripteurs SIFT caractérisant le contenu visuel d'une image, nous avons utilisé la bibliothèque "open-source SIFT" qui est disponible à <http://eecs.oregonstate.edu/~hess/sift.html>. Cette bibliothèque est écrite en C, avec des versions disponibles à la fois pour Linux et Windows. En particulier les fonctions de la bibliothèque SIFT utilisent des types de données OpenCV pour représenter les images, les matrices, etc., ce qui facilite d'effectuer des opérations interne de la bibliothèque SIFT en utilisant les fonctions d'OpenCV.

La bibliothèque SIFT contient quatre composants principaux, chacun étant présenté par un fichier d'entête différent:

### ✓ *Détection des points clés*

Le principal composant de la bibliothèque SIFT est un ensemble de fonctions qui permet de détecter les points clés SIFT. La bibliothèque contient deux fonctions de détection des points clés (situé dans le fichier d'entête sift.h), l'une des fonctions calcule les points clés en utilisant des paramètres par défaut proposé dans le document de Lowe [73], et une autre fonction qui donne le choix à l'utilisateur pour définir les paramètres. Ces fonctions sont conçues pour être faciles d'appeler. Spécifiquement, elles ne nécessitent pas des appels aux

fonctions d'initialisation, et elle permet d'accepter à la fois des images en niveau de gris et des images RGB.

```
IplImage *img; /*type d'image OpenCV*/
Structfeature *keypoints; /*SIFT*/
int n; /*nombre des caractéristiques*/

/*lire une image en utilisant OpenCV et détecter les
    poits clés */
img = cvLoadImage("/path/to/image.png",1);
n = sift_features(img,&keypoints);
```

### ✓ *Apprentissage des Kd-tree*

La bibliothèque SIFT contient des structures et des fonctions (situé dans le fichier d'entête `kdtree.h`), qui permet la mise en correspondance entre des points clés calculés à partir d'une image données et les points clés calculée à partir d'une image de la base de données.

```
/* trouver l'image la plus proche dans kd-tree en
    utilisant la stratégie best bin first*/
/*Paramètres*/

@param kd_root root of an image feature kd tree
@param feat image feature for whose neighbors to
    search
@param k number of neighbors to find
@param nbrs pointer to an array in which to store
    pointers to neighbors
    in order of increasing descriptor distance
@param max_nn_chks search is cut off after
    examining this many tree entries

@return Returns le nombre des points proche trouvés
    et sauvegarder dans nbrs, ou bien
    -1 en cas d'erreur.
/*Fonction*/
int kdtree_bbf_knn( struct kd_node* kd_root, struct
    feature* feat, int k, struct feature*** nbrs, int
    max_nn_chks )
```

✓ *Calcul de la transformation RANSAC*

La bibliothèque SIFT contient des structures et des fonctions (situé dans le fichier d'entête `xform.h`) qui permet de calculer la matrice de transformation à partir des correspondances entre les points clés. Une fois les correspondances entre les caractéristiques des images sont réalisées, l'algorithme RANSAC est utilisé pour effectuer le calcul de la matrice de transformation.

✓ *Manipulation des caractéristiques d'images invariantes*

En fin la bibliothèque SIFT contient un ensemble de structures et de fonctions (situé dans le fichier d'entête `imgfeatures.h`) qui permet de manipuler des caractéristiques d'images invariants. En particulier cette composante de la bibliothèque contient des fonctions qui permet de lire les caractéristiques d'une image à partir d'un fichier, à condition que ce fichier contient des points clé calculés en utilisant les fonctions de la bibliothèque SIFT, ou bien des caractéristiques calculé par Oxford Visual Geometry Group's software. Cette bibliothèque contient aussi des fonctions qui permettent d'exporter les caractéristiques calculées dans un autre fichier.

La bibliothèque SIFT contient aussi trois exemples simples d'application décrits ci-dessous.

- **Siftfeat.c:** cette application calcule les points clé SIFT d'une image et les mettre dans un fichier.
- **Match.c:** cette application détecte les points clés de deux images puis elle trouve les correspondances entre eux en utilisant les fonctions kd-tree de la bibliothèque, et calcule éventuellement une transformé sur la base de ces correspondances en utilisant les fonctions RANSAC de la bibliothèque.
- **Dspfeat.c:** cette application affiche les caractéristiques des images importées à partir d'un fichier.

## 2.2. SVM

Pour l'apprentissage et la classification on a utilisé Libsvm un programme simple, facile à utiliser, et efficaces pour SVM classification et régression. Libsvm est disponible à <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

Le format des fichiers de données d'apprentissage et de test est la suivante:

```
<label> <index1>: <valeur1> <index2>: <valeur2> ...  
.  
.  
.
```

Chaque ligne contient une instance et se termine par un '\ n'. Pour la classification, <label> est un entier indiquant l'étiquette de classe. <index>:<valeur> donne la valeur d'une caractéristique. <index> est nombre entier débute de 1 et <valeur> est nombre réel. Labels dans le fichier test ne sont utilisés que pour calculer les précisions ou des erreurs. Si elles sont inconnues, il suffit de remplir la première colonne avec n'importe quel numéro.

La structure `svm_problem` décrit le problème:

```
struct svm_problem  
{  
    int l;  
    double *y;  
    struct svm_node **x;  
};
```

où «L» est le nombre de données d'apprentissage, et "y" est un tableau contenant leurs classes. «x» est un tableau de pointeurs, dont chacun des points est tableau de `svm_node` d'un vecteur d'apprentissage.

```
struct svm_node  
{  
    int index;  
    double value;  
};
```

index = -1 indique la fin d'un vecteur. Notez que les indices doivent être dans l'ordre croissant. `struct svm_parameter` décrit les paramètres d'un modèle SVM:

```

struct svm_parameter
{
    int svm_type;
    int kernel_type;
    int degree; /* pour poly */
    double gamma; /* pour poly/rbf/sigmoid */
    double coef0; /* pour poly/sigmoid */

    /* pour l'apprentissage seulement */
    double cache_size; /* en MB */
    double eps; /* critères d'arrêt */
    double C; /* pour C_SVC, EPSILON_SVR, and NU_SVR */
    int nr_weight; /* pour C_SVC */
    int *weight_label; /* pour C_SVC */
    double* weight; /* pour C_SVC */
    double nu; /* pour NU_SVC, ONE_CLASS, and NU_SVR */
    double p; /* pour EPSILON_SVR */
    int shrinking;
    int probability; /* faire des estimations de probabilité */
};

```

struct svm\_model décrit le modèle généré par un SVM:

```

struct svm_model
{
    struct svm_parameter param; /* paramètres */
    int nr_class; /* nombre de classes */
    int l; /* totale #SV */
    struct svm_node **SV; /* SVs (SV[l]) */
    double **sv_coef; /* coefficients pour SVs dans les
fonctions de decision (sv_coef[k-1][l]) */
    double *rho; /* constants des fonctions de decision
(rho[k*(k-1)/2]) */
    double *probA;
    double *probB;

    /* pour la classification seulement */

    int *label; /* label de chaque classe (label[k]) */
    int *nSV; /* nombre de SVs pour chaque classe (nSV[k]) */
/* nSV[0] + nSV[1] + ... + nSV[k-1] = l */
    int free_sv; /* 1 si svm_model est créée par
svm_load_model*/
/* 0 si svm_model est créée par svm_train */
};

```

```
- Function: struct svm_model *svm_train(const struct svm_problem *prob, const struct svm_parameter *param);
```

Cette fonction construit et retourne un modèle SVM selon les données d'apprentissages `prob` et les paramètres `param`.

```
- Function: double svm_predict(const struct svm_model *model, const struct svm_node *x);
```

Etant donné un modèle `model` et un vecteur de test `x`, cette fonction retourne la classe prévue pour `x`.

### 3. Tests et Expérimentations

#### 3.1. Base de données

Pour valider notre système nous avons utilisé la base de données ORL [78], cette base contient un ensemble des images prise entre avril 1992 et avril 1994 au laboratoire de recherche Olivetti à Cambridge. Il y a 10 images différentes pour 40 personnes distinctes (taille total de la base de données: 400 images). Les visages présente des expressions faciales (sourire / non sourire, les yeux ouverts / fermés) et des détails faciaux (lunettes / pas de lunettes). Les images ont été prise à temps différentes, induisant une variation supplémentaire inhérente à l'âge. La taille de chaque image est 92X112 pixels, chaque pixel est codé sur un octet. Afin de tester et évaluer notre système, nous avons défini trois sous bases de données visages pour l'apprentissage (A1, A2, A3) associées respectivement à trois sous bases de données visages pour le test (T1, T2, T3)

#### ❖ Bases d'apprentissage:

**A1:** Les deux premières images servent pour la phase d'apprentissage, ce qui nous fait un sous ensemble de 80 images dédiées à cette phase.

**A2:** Les cinq premières images servent pour la phase d'apprentissage, ce qui nous fait un sous ensemble de 200 images dédiées à cette phase.

**A1:** Les huit premières images servent pour la phase d'apprentissage, ce qui nous fait un sous ensemble de 320 images dédiées à cette phase.

❖ **Base de test:**

**T1:** Les huit dernières images de chaque individu nous ont servies pour la réalisation des différents tests, ce qui nous fait un sous ensemble de 320 images dédiées à la phase test

**T2:** Les cinq dernières images de chaque individu nous ont servies pour la réalisation des différents tests, ce qui nous fait un sous ensemble de 200 images dédiées à la phase test

**T3:** Les deux dernières images de chaque individu nous ont servies pour la réalisation des différents tests, ce qui nous fait un sous ensemble de 80 images dédiées à la phase test.



**FIG31 :** Des exemples de la base de données ORL

Les images de la base de données ORL sont utilisées sans aucune type de prétraitement, et cela pour évaluer la robustesse des algorithmes dans la comparaison.

### 3.2. SIFT

On a une moyenne de 70 caractéristiques SIFT extraites de chaque image de la base de données.



FIG32: 3 exemples d'images



FIG33: les images avec les descripteurs SIFT

La nouvelle approche SIFT est comparée avec la méthode Eigenfaces. Le tableau 3 montre les résultats obtenus. Les performances se dégradent avec l'ensemble d'apprentissage de plus petite taille et augmente avec l'ensemble d'apprentissage plus large. Il est également clair que l'approche SIFT est meilleure que l'approche Eigenfaces. La performance est nettement meilleure en utilisant 80% des données pour l'apprentissage et seulement 20% des données pour le test (99,3% pour les SIFT contre 97,2% pour Eigenfaces).

<i>Bases</i>	<i>Méthode</i>	SIFT	Eigenfaces
A1,T1		99,3%	97,2%
A2,T2		96,3%	92,9%
A3,T3		85,7%	80,1%

Tableau 3 : Comparaison entre l'approche SIFT et Eigenfaces

### 3.3. SVM

Avec le classificateur SVM, plusieurs noyaux peuvent être utilisés pour faire l'apprentissage et à chaque noyau correspondent différents paramètres à faire varier. L'utilisateur sera contraint de définir des paramètres qui vont être utilisés dans l'apprentissage:

- ✓ Choix du noyau : L'utilisateur choisit le noyau à utiliser pour la construction du classificateur. Les noyaux les plus populaires sont : le noyau linéaire, le noyau polynomial et le noyau gaussien.
- ✓ Choix de paramètres général C et des paramètres spécifiques pour chaque noyau. Le noyau linéaire n'a pas de paramètres spécifiques, tandis que si on choisit le noyau gaussien il faut fixer le paramètre sigma. Quand au noyau polynomial, il faut fixer le degré du noyau polynomial.

Nous allons exposer les différents tableaux et graphes obtenus lors de l'évaluation de notre système en testant tous les types de noyaux.

#### 3.3.1. Noyau linéaire

Cette expérience est testée en faisant varier la constante C du classifieur SVM. C est un hyper-paramètre qui régit la performance du SVM. Ce paramètre sert à fixer le compromis entre la minimisation de l'erreur d'apprentissage et la maximisation de la marge. En pratique, le comportement du SVM est sensible à la valeur de C uniquement si les données d'apprentissage ne sont pas séparables. Une très grande valeur de C peut faire que la fonction objective minimisée par le SVM ne soit plus convexe et empêcherait sa convergence. Une très faible valeur de C tend à diminuer la capacité du classifieur.

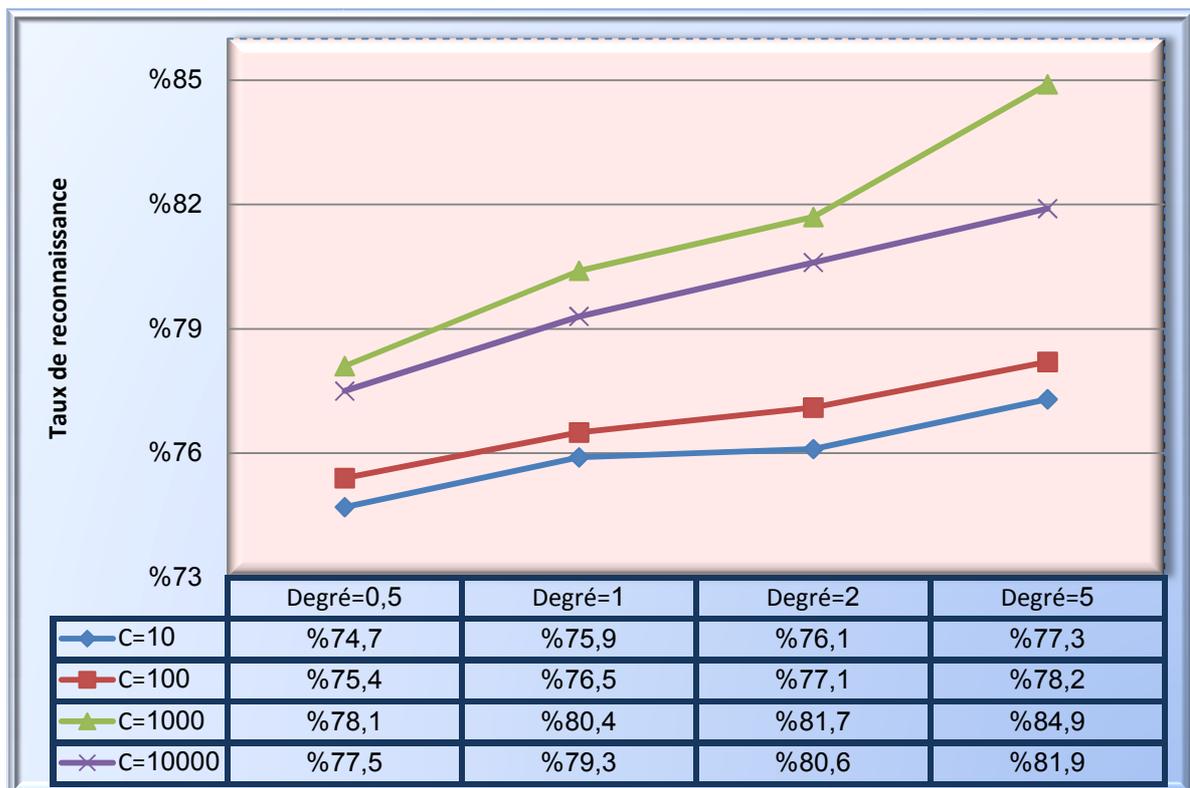
Noyau Linéaire	C=10	C=100	C=1000	C=10000
A1,T1	37, 3%	41,5%	49,1%	43,2%
A2,T2	51,1%	53, 4%	56,8%	52,7%
A3,T3	53,4%	55,7%	60,4%	56,9%

**Tableau 4** : les résultats obtenus en utilisant le noyau linéaire

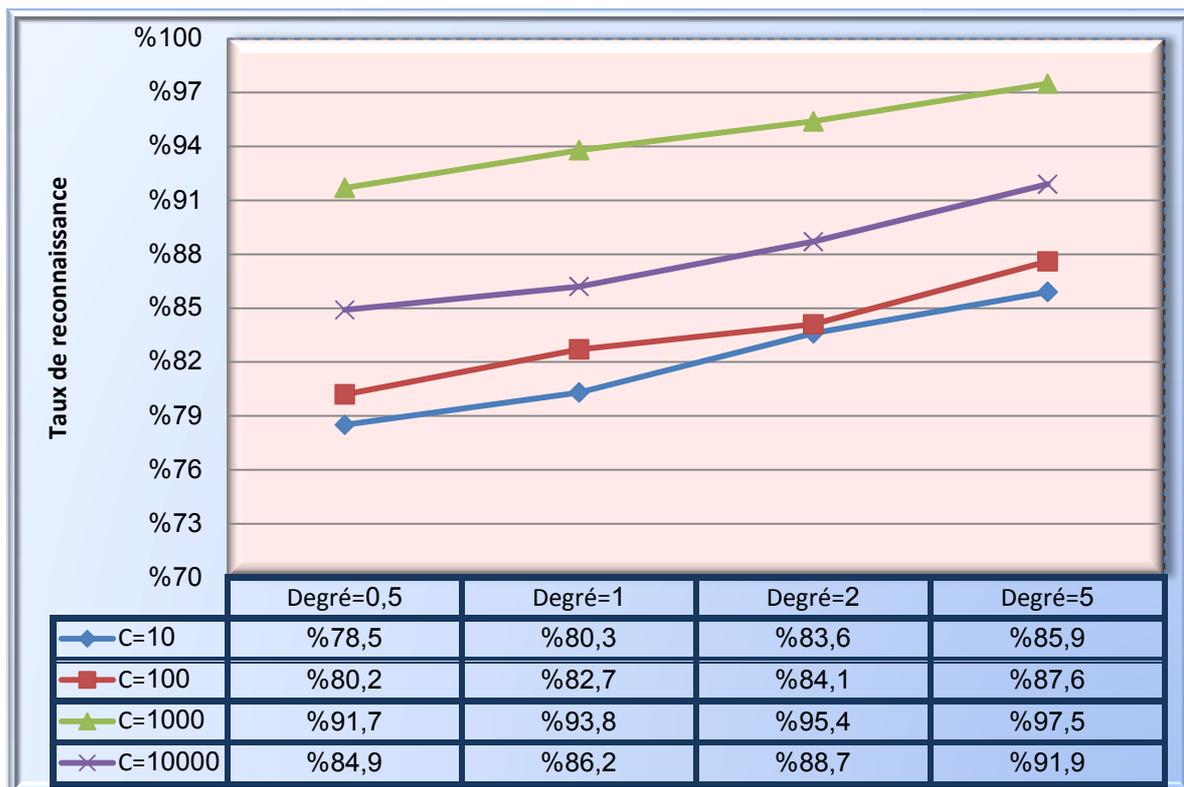
Les résultats obtenus en utilisant le noyau linéaire montrent la faiblesse de ce noyau qui donne un taux de reconnaissance bas. Cela peut être expliqué par le fait que les classes des visages sont non linéairement séparables, ce qui implique la présence de données mal classées.

### 3.3.2. Noyau polynomial

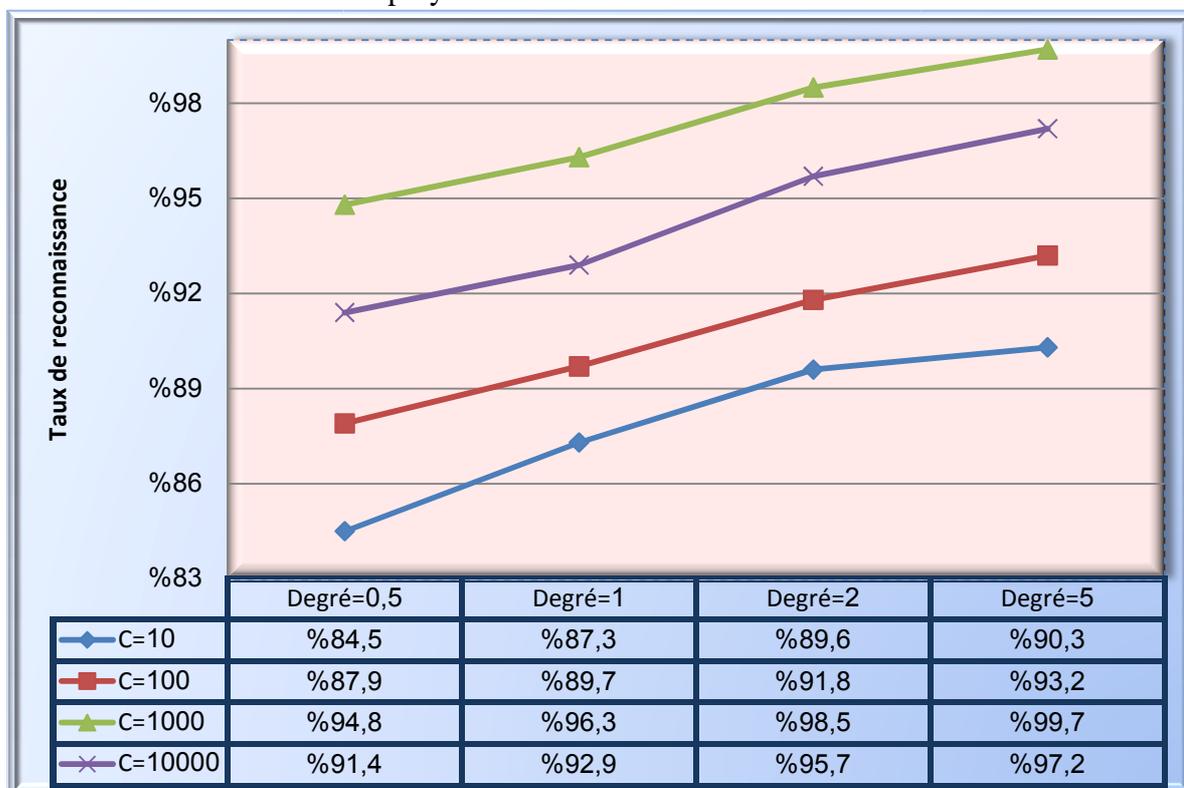
On utilisant le noyau polynomial on cherche à trouver une séparation linéaire dans un espace à haute dimension. La nouvelle dimension est en relation direct avec le degré du noyau polynomial. Les données vont être séparée par un hyperplan dans un espace à 3 dimensions si on utilise un noyau polynomial de degré 2, et ça sera un espace à 4 dimensions si on utilise un polynôme de degré 3. Les figures 34, 35, 36 donnent les taux de reconnaissance du noyau polynomial, en faisant varier: le degré du polynôme de la valeur 2 jusqu'au 5, et la taille de la base d'apprentissage et de test. Les résultats augmentent au fur et à mesure que la valeur de degré augmente, ils sont meilleurs pour la valeur 5 du degré et pour C=1000.



**FIG34:** courbes représentant le taux de reconnaissance du noyau polynomial en utilisant A1 et T1



**FIG35:** courbes représentant le taux de reconnaissance du noyau polynomial en utilisant A2 et T2



**FIG36:** courbes représentant le taux de reconnaissance du noyau polynomial en utilisant A3 et T3

L'utilisation du noyau polynomial qui est un noyau non linéaire montre une nette amélioration des performances par rapport au noyau linéaire (99,7% en utilisant 80% des données pour l'apprentissage et 20% des données pour le test).

### 3.3.3. Noyau RBF

En utilisant le noyau RBF, on cherche à trouver une séparation linéaire dans un espace à haute dimension. Au contraire au noyau polynomial la nouvelle dimension obtenue en utilisant le noyau RBF est inconnu et elle tend vers l'infini. Le noyau RBF est caractérisé par la variable sigma. Elle représente les degrés de similarité entre les exemples d'apprentissage. Si elle est très petite on risque de tomber dans le problème de sur apprentissage. Les figures 37, 38, 39 illustre l'effet de la variable sigma sur la reconnaissance:

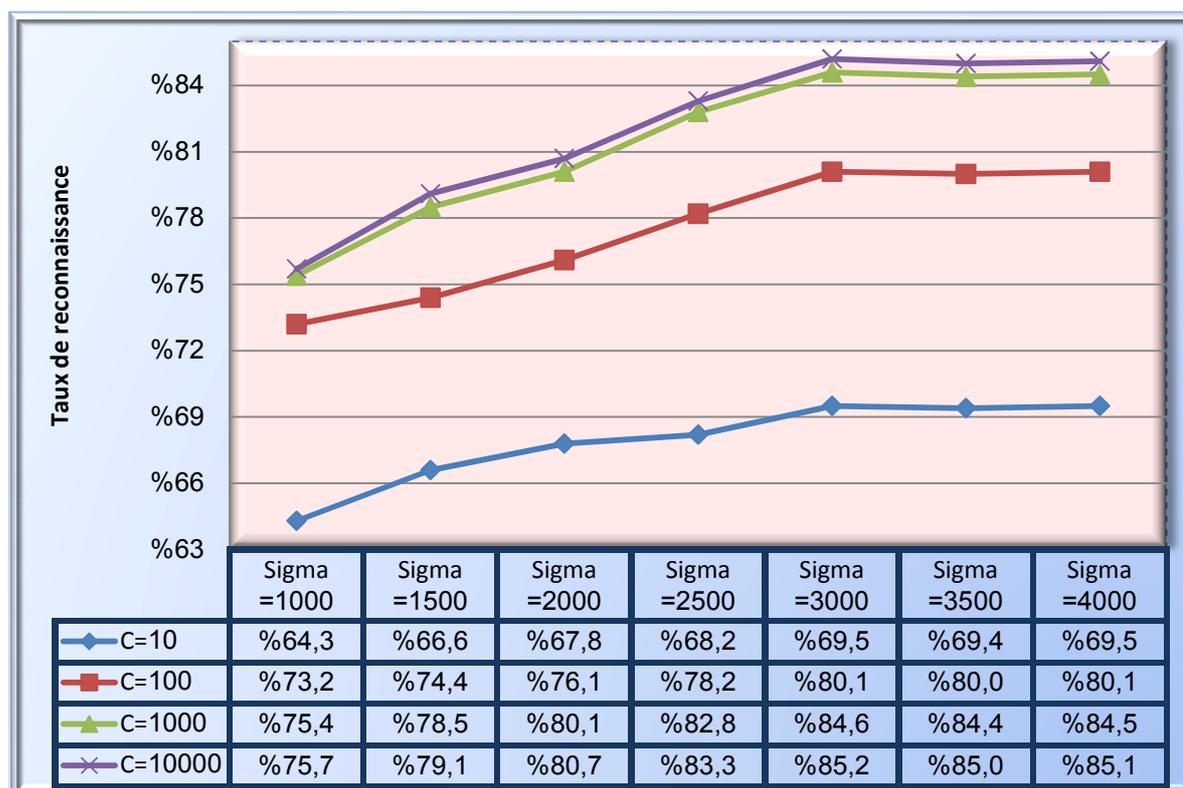


FIG37: courbes représentant le taux de reconnaissance du noyau RBF en utilisant A1 et T1

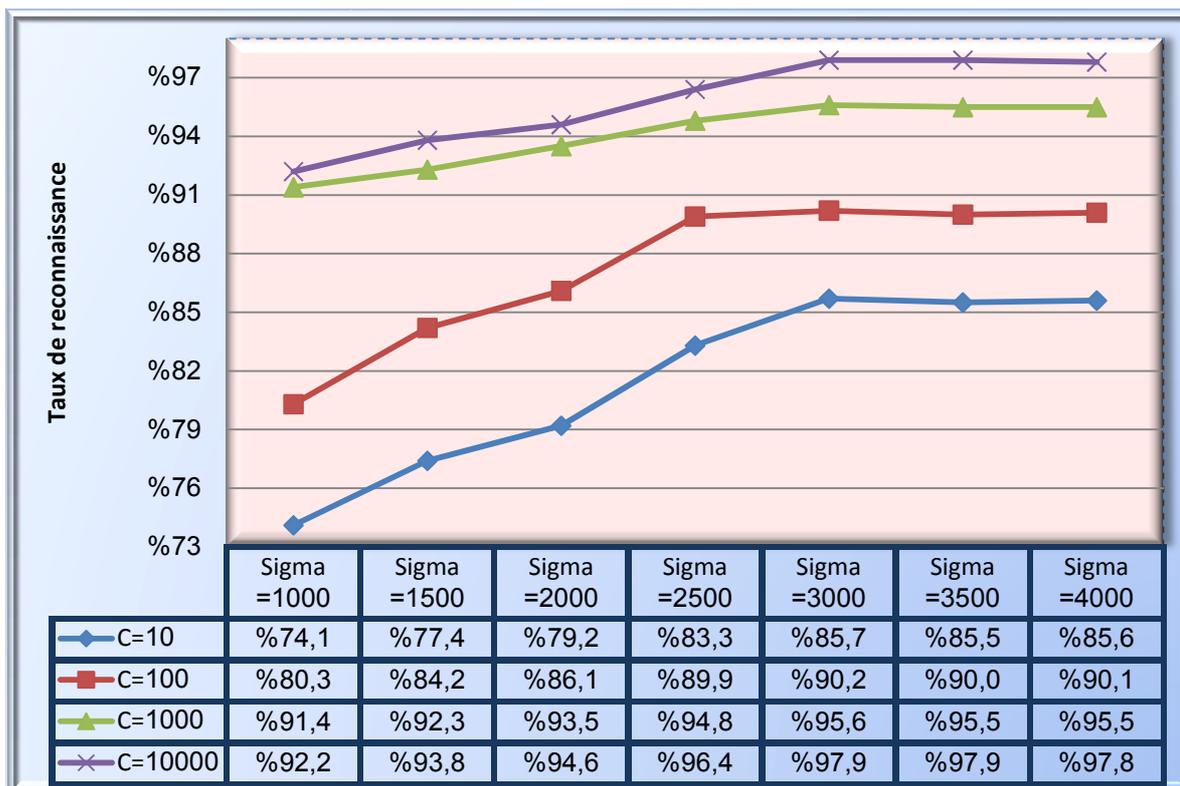


FIG38: courbes représentant le taux de reconnaissance du noyau RBF en utilisant A2 et T2

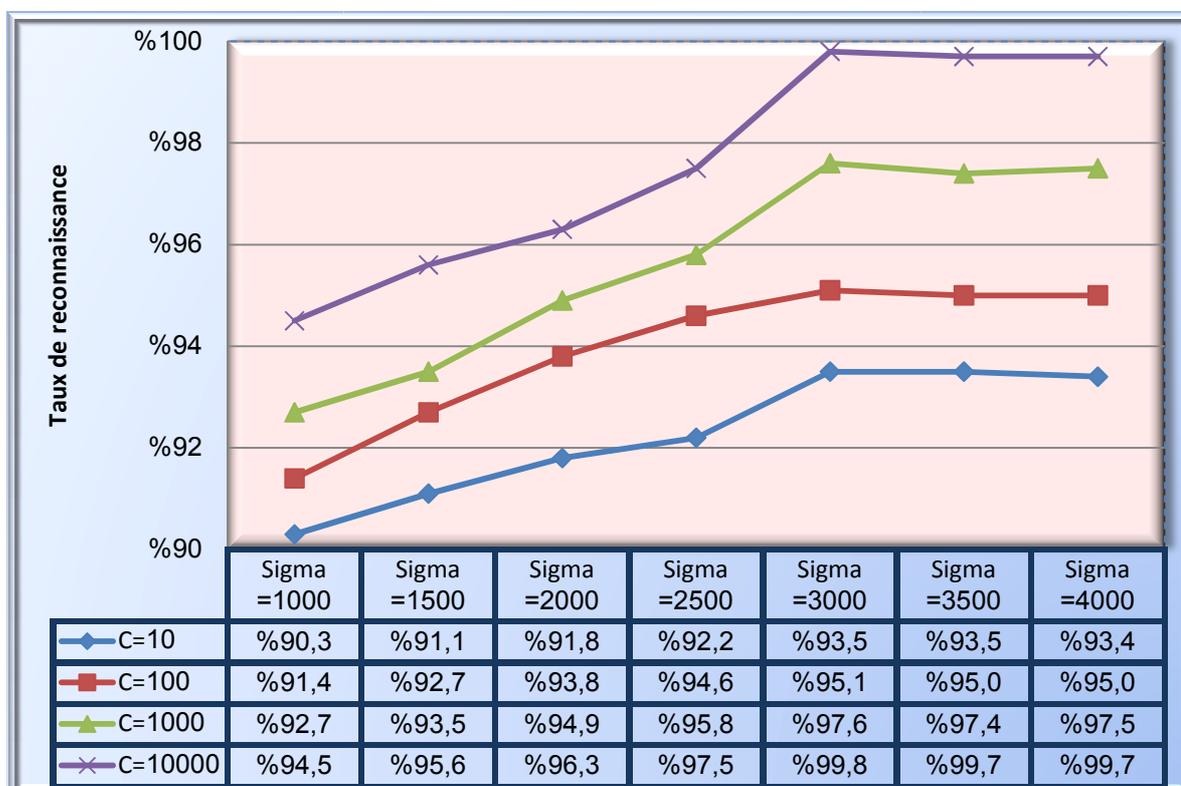


FIG39: courbes représentant le taux de reconnaissance du noyau RBF en utilisant A3 et T3

Le noyau RBF qui est aussi non linéaire améliore la reconnaissance. Les courbes représentées sur la figure 38 montrent que l'augmentation du sigma jusqu'à la valeur 3000 augmente le taux de reconnaissance. A partir de cette valeur, le taux de reconnaissance devient presque fixe. Appliquer un sigma très petit engendre un sur-apprentissage (overfitting) des données, c'est-à-dire que la fonction de décision colle les données d'apprentissage et cela détériore le pouvoir de généralisation du classificateur, ce qui entraîne des erreurs de mauvaise classification des nouvelles données. En revanche, un apprentissage avec un sigma très grand, nous fait tomber dans le cas du sous-apprentissage, c'est-à-dire, le cas d'une fonction de décision séparant les classes avec une erreur empirique élevée.

Les meilleurs taux de reconnaissance avec la méthode SVM sont représentés dans le tableau 5 :

	Noyau linéaire	Noyau polynomial	Noyau RBF
Paramètres	<b>C=1000</b>	<b>C=1000 et degré=5</b>	<b>C=10000 et sigma=3000</b>
Taux de reconnaissance	<b>60,4%</b>	<b>99,7%</b>	<b>99,8%</b>

**Tableau 5** : les meilleurs résultats obtenus pour chaque noyau

Pour différents noyaux du classifieur SVM et pour différentes valeurs de C, de Degré du polynôme et de Sigma les résultats obtenus dénotent que le noyau RBF est le meilleur ; il permet d'obtenir des taux de classification très importants.

Méthode	SIFT	SVM
<b>Bases</b>		
<b>A1,T1</b>	<b>99,3%</b>	<b>99,8%</b>
<b>A2,T2</b>	<b>96,3%</b>	<b>97,9%</b>
<b>A3,T3</b>	<b>85,7%</b>	<b>85,2%</b>

**Tableau6** : les meilleurs résultats obtenus pour chaque méthode

#### 4. Bilan global et conclusion

Dans ce chapitre nous avons présenté une série d'expérimentations en utilisant les méthodes SIFT et SVM. Les performances se dégradent avec l'ensemble d'apprentissage de plus petite taille et augmente avec l'ensemble d'apprentissage plus large. Les résultats montrent que l'approche SIFT est meilleure par rapport à l'approche Eigenfaces. Des expériences sont faites en utilisant le noyau linéaire, polynomiale et RBF. Le noyau linéaire est testée en faisant varier le paramètre  $C$ ; une très grande valeur de  $C$  peut faire que la fonction objective minimisée par le SVM ne soit plus convexe et empêcherait sa convergence; une très faible valeur de  $C$  tend à diminuer la capacité du classifieur. Les résultats obtenus en utilisant le noyau linéaire montrent la faiblesse de ce noyau qui donne un taux de reconnaissance bas. L'utilisation du noyau polynomial en faisant varier le degré du polynôme montre une nette amélioration des performances par rapport au noyau linéaire. Le noyau RBF qui est caractérisé par le degré de similarité  $\sigma$  améliore la reconnaissance; appliquer un  $\sigma$  très petit engendre un sur-apprentissage (overfitting) des données; en revanche, un apprentissage avec un  $\sigma$  très grand, nous fait tomber dans le cas du sous-apprentissage.

Les résultats de la classification par SVM obtenus peuvent être améliorés en optimisant les paramètres suivants:

- ✓ Les paramètres  $C$ , Degré du polynôme et  $\sigma$ .
- ✓ Le nombre d'exemple d'apprentissage

## *Conclusion et perspectives*

Ce travail s'inscrit dans le domaine de la reconnaissance automatique des visages. Celle-ci consiste à vérifier l'identité d'une personne à partir de son image. Utilisés principalement pour des raisons de sécurité et/ou de confidentialité, les systèmes de reconnaissance automatique des visages sont souvent développés dans les applications de télésurveillance et l'accès à des endroits sécurisés.

Dans ce mémoire, nous avons traité deux problématiques majeures et complémentaires rencontrées en reconnaissance de visage. Il s'agit, d'une part de l'extraction automatique des caractéristiques de visage, et d'autre part de la reconnaissance du visage.

La difficulté de l'extraction des caractéristiques de visage est due principalement aux variations des poses, des expressions faciales et des conditions d'éclairage. Nous avons donc proposé d'utiliser les descripteurs SIFT qui caractérisent le contenu visuel de l'image d'une façon indépendante de l'échelle ("zoom" et résolution du capteur), du cadrage, de l'angle d'observation et de l'exposition (luminosité). L'algorithme SIFT procède en quatre étapes principales : la détection des extremums (qui correspondent aux points d'intérêt) à un espace d'échelle créé par des filtre DoG sur une pyramide multi-échelle; la localisation et le filtrage des points d'intérêt; le calcul des histogrammes des directions des gradients locaux autour du point d'intérêt dans un voisinage qui varie en fonction de l'échelle du point, et le calcul d'une signature à partir des histogrammes en divisant la zone autour du point en quatre sous-zones.

Pour l'apprentissage et la reconnaissance nous avons utilisé la méthode SVM. Les trois types des noyaux: linéaire, polynomial et RBF, sont vérifiés avec différents paramètres. Les meilleurs résultats (taux de reconnaissance = 99,8%) sont obtenus en utilisant le noyau RBF avec la valeur 3000 de sigma et la valeur 10000 de C.

Comme perspective de ce travail nous souhaitons améliorer le module d'extraction automatique des caractéristiques de visage, nous pensons que l'utilisation des informations 3D améliorerait la robustesse de la méthode d'extraction des caractéristiques.

# Annexe A

## Les bases de données utilisées

Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance de visages sont disponibles sur le marché. Toutefois, ces bases de données sont généralement adaptées aux besoins de quelques algorithmes spécifiques de reconnaissance. Dans cette section, nous décrivons les bases les plus utilisées. Le tableau 7 récapitule les principales caractéristiques de ces bases. Il résume, pour chacune d'elles, les conditions d'acquisition des images de visages (changements d'illumination, de pose, d'expression faciale, occultation et variation dans le temps), "-" : dénote le cas où l'un de ces éléments n'a pas été mesuré, ou était non contrôlé durant la prise de vue.

Base	Nombre de personnes	Pose	Illumination	Expression	Occultation	Nombre de sessions
<b>FERET</b>	1199	1-9	1-2	2	-	2
<b>Yale</b>	15	3	3	6	1	1
<b>ORL</b>	40	-	-	-	-	-
<b>PF01</b>	107	8	4	5	-	1
<b>UMIST</b>	20	-	-	-	-	1
<b>AR</b>	116	1	4	4	2	2
<b>PIE</b>	68	13	43	4	-	1

**Tableau7** : Principales Caractéristiques des bases de visages [75].

## 1. La base FERET

La base FERET a été collectée dans le cadre du programme Facial Recognition Technology [76] mené par National Institute of Standards and Technology (NIST) Américain. Il s'agit à ce jour de plus grande base disponible gratuitement pour les chercheurs. Les images, initialement collectées depuis un appareil photographique de 35mm, ont ensuite été digitalisées. Un extrait de cette base est donnée en figure 37. Elle contient plus de 14051 images de résolution 256 x 384, représentant 1199 personnes. Les images ont été collectées lors de 15 sessions entre Août 1993 et Juillet 1996. Pour chaque individu, on dispose de deux vues (fa et fb) avec des expressions faciales différentes (généralement, une expression neutre et deux sourie). Pour 200 de ces personnes, on dispose d'une troisième image prise avec une caméra et des conditions d'illumination différentes (vue fc). Pour ces 200 individus, la base contient des vues additionnels montrant des changements de pose en profondeur (allant du profil gauche au profil droit). Pour quelques personnes, on dispose de deux autres vues collectées dans des conditions similaires à fa et fb mais à des dates différentes (vue duplicate). Aucune contrainte n'est imposée sur la date de la prise de vue de l'image duplicate I. par contre, la vue duplicate II a été collectée au moins un an après la première prise de vue.

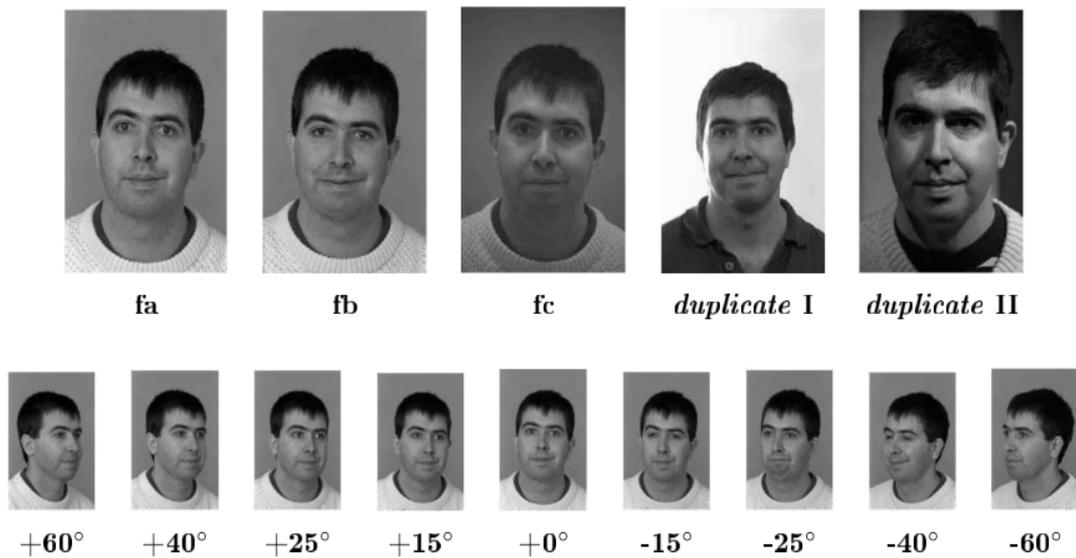


FIG40 : Extrait de la base FERET

## 2. La base de Yale

La base de visage de Yale [77], illustrée en figure 38, contient 165 vues de 15 personnes (11 vues par personne). La taille des images est de 320 x 243 pixels. Pour chaque personne, on dispose de huit vues sous conditions d'illumination neutres, avec: une expression faciale neutre, un sourire, le clignent d'un œil, les deux yeux fermés, une expression de surprise, une expression triste et une vue neutre avec et sans lunettes. Les trois vues restantes correspondent à des changements d'illumination: l'une avec une illumination de face, la seconde avec une illumination provenant du côté droit, la troisième venant du côté gauche. Toute les vues d'une personne ne portant habituellement pas de lunettes (première ligne de la figure 39 ) sont représentées sans lunettes sauf une pour deux sujets (exemple en deuxième ligne de la figure 39), on dispose de dix vues avec lunettes et d'une seule vue sans.

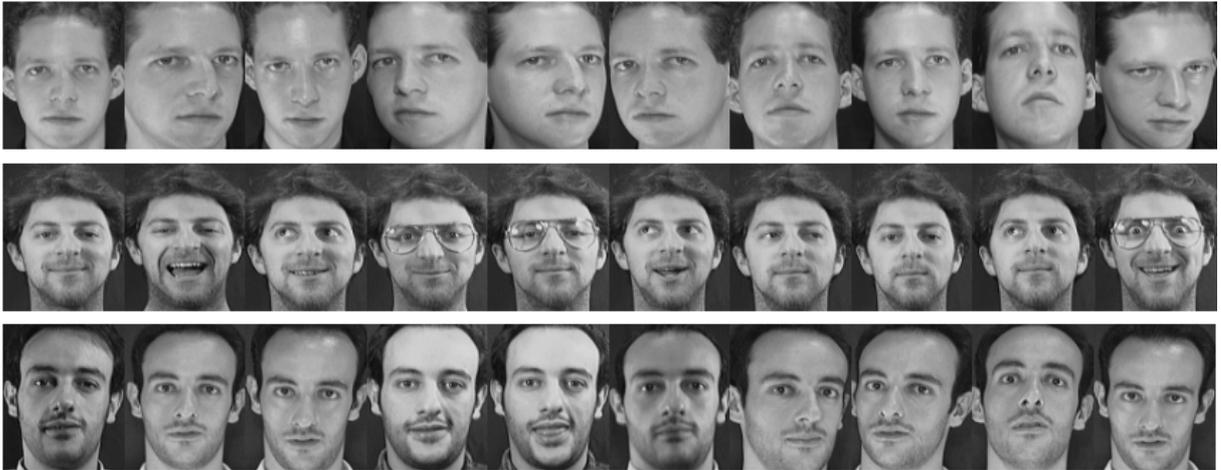


FIG41 : Extrait de la base de Yale

## 3. La base ORL

La base ORL [78] a été collectée dans le cadre d'un projet mené par un laboratoire de AT&T, basé à Cambridge, en collaboration avec l'université de Cambridge. Les prises de vue ont été menées entre Avril 1992 et Avril 1994. La base contient 40 personnes, chacune étant enregistrée sous 10 vues différentes (figure 40). Les images sont de taille 112 x 92 pixels. Pour quelques sujets, les images ont été collectées à des dates différentes, avec des variations dans les conditions d'illumination, les expressions faciales (expression neutre, sourire et yeux fermés) et des occultations partielles (port de lunette ou non). Toutes les images ont été collectées sur un fond foncé. Les poses de la tête présentent quelques variations en profondeur par rapport à la pose frontale. Cette base fait partie de celles qui ont été le plus utilisées et permet de comparer facilement les performances de tout nouvel algorithme. Etant donné que les variations ne portent que sur certaines

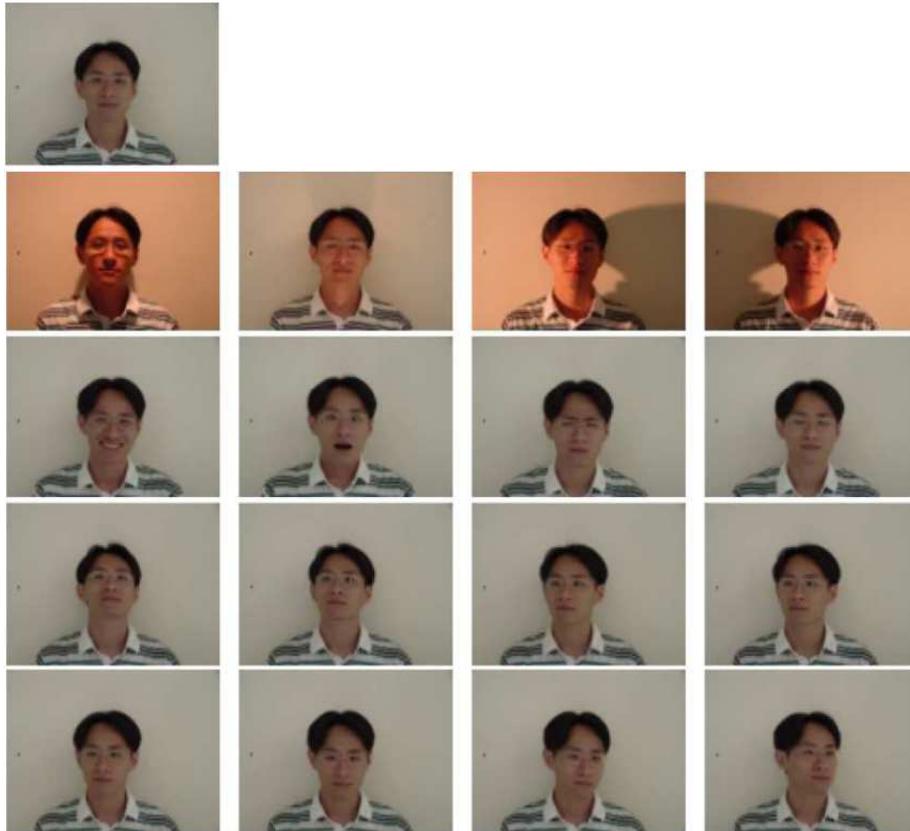
personnes et ne sont donc pas systématiques, cette base ne peut cependant pas être utilisée pour mener une analyse de sensibilité à différents facteurs.



**FIG42:** Extrait de la base ORL

#### 4. La base PF01

La base PF01 (Postech Faces 01) [79] a été développée par le laboratoire Intelligent Multimedia Laboratory (IML) de l'université de Pohang, en Corée. Elle contient des vues de 107 personnes d'origine asiatique (56 hommes et 51 femmes). Une trentaine d'entre elles portent des lunettes. Pour chaque personne, on dispose de 17 vues, comme le montre la figure 40. Les images sont en couleurs et de taille 1280 x 960 pixels. Treize vues par personne sont prises dans des conditions d'illumination neutres et standardisées. Parmi ces treize vues il y en a une qui représente une expression faciale neutre, quatre qui correspondent à des changements d'expression faciale et huit qui montrent des changements de pose en profondeur. Les quatre vues restantes, collectées sous une pose frontale et avec une expression faciale neutre, contiennent des changements d'illumination. Cette base, disponible gratuitement sur internet sur le site d'IML <http://nova.postech.ac.kr/>, est très intéressante pour étudier les effets de changements de pose et d'expression faciale. Elle présente de plus l'avantage de compter de nombreuses vues pour un nombre de personnes enregistrées de l'ordre de la centaine.



**FIG43** : Extrait de la base PF01

## 5. La base UMIST

La base UMIST [80] contient 564 vues de 20 personnes. Les images sont en niveaux de gris et de résolution 220 X 220. Elles sont tirées de séquences durant lesquelles les sujets tournent lentement la tête du profil à une vue frontale (figure42). Les sujets représentés sont de différents sexes, âges et origines ethnique.



FIG44 : Extrait de la base UMIST

## 6. La base AR

La base AR [81] a été constituée en 1998 sein du laboratoire Computer Vision Center (CVC) à Barcelone, en Espagne. 116 personnes (63 hommes et 53 femmes) sont enregistrées. Les images sont en couleur, de taille 768 x 576 pixels. 26 vues de chacun de ces sujets ont été collectées lors de deux sessions, menées à deux semaines d'intervalle. Lors de chaque session, 13 vues par personne ont été enregistrées. Un extrait des images

collectées lors de la première session est fourni en figure 43. Ces vues présentent des changements d'expression faciale, d'illumination, ainsi que des occultations partielles des yeux (port de lunettes) et de la partie basse du visage (cache-nez). Lors de la seconde session, les 13 vues sont collectées dans les mêmes conditions que pour la première session.

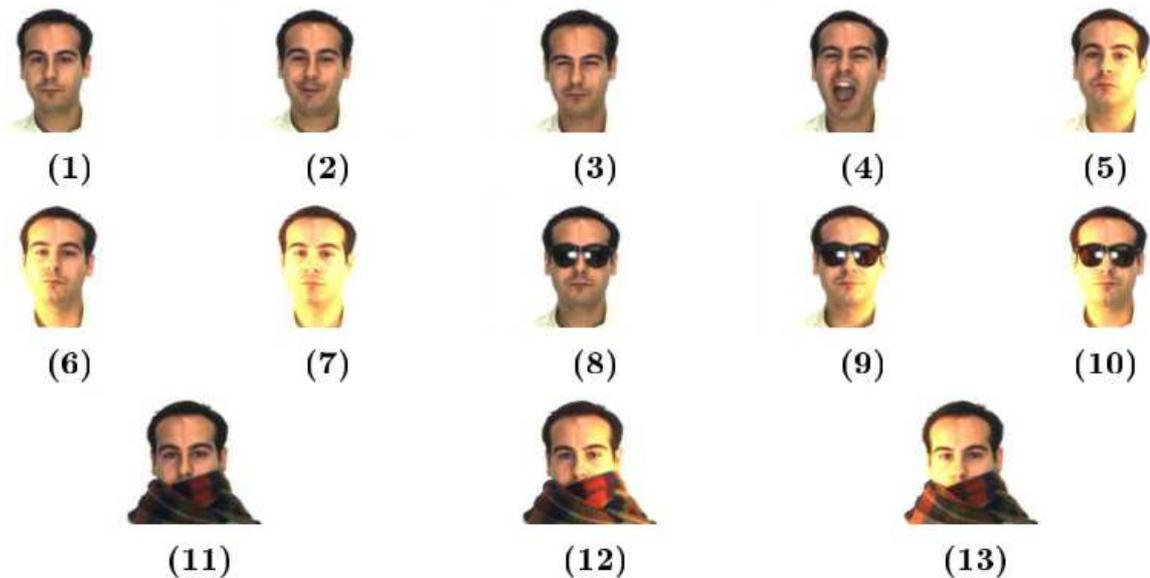


FIG45: Extrait de la base AR

## 7. La base PIE

La PIE [82] a été collectée au sein de la Carnegie Mellon University entre Octobre et Décembre 2000. La base contient 41368 images de 68 personnes. Les images sont en couleur et de taille 640 x 480 pixels. Les angles de prise de vue, les conditions d'illumination ainsi que les expressions faciales sont systématiquement variées. Les vues ont été collectées dans une salle spécifique (CMU 3D) équipée de 13 appareils photographiques synchronisés de haute qualité et de 21 flashes. La figure 44 montre les vues d'un sujet sous les 13 angles de prise de vue différents. En plus des variations de pose, quatre autres facteurs ont été pris en compte:

- **Illumination I** (voir figure 44 gauche): les 21 flashes sont activés séquentiellement de manière très rapide. Les lumières de la pièce sont allumées;

- **Illumination II** (voir figure 44 droite): les 21 flashes sont activés de la même manière que pour l'illumination I, mais avec les lumières de la pièce éteintes;
- **Expression**: les expressions représentées sont: expression neutre, sourire, clignement d'un œil et les deux yeux fermés. Les images collectées par les 13 caméras sont disponibles dans la base;
- **Parole**: 60 vues de chaque personne en train de parler sont enregistrées sous trois angles de prise de vue différents (de face, de trois-quarts et de profil).

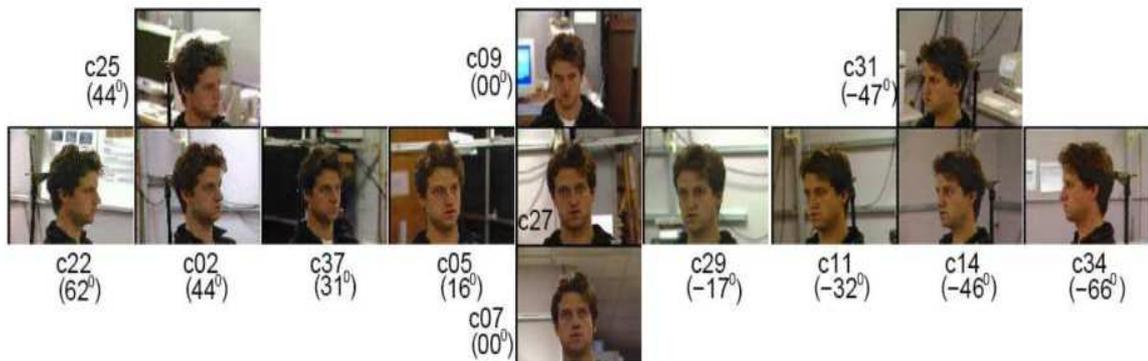


FIG46: Extrait de la base PIE: variation de pose

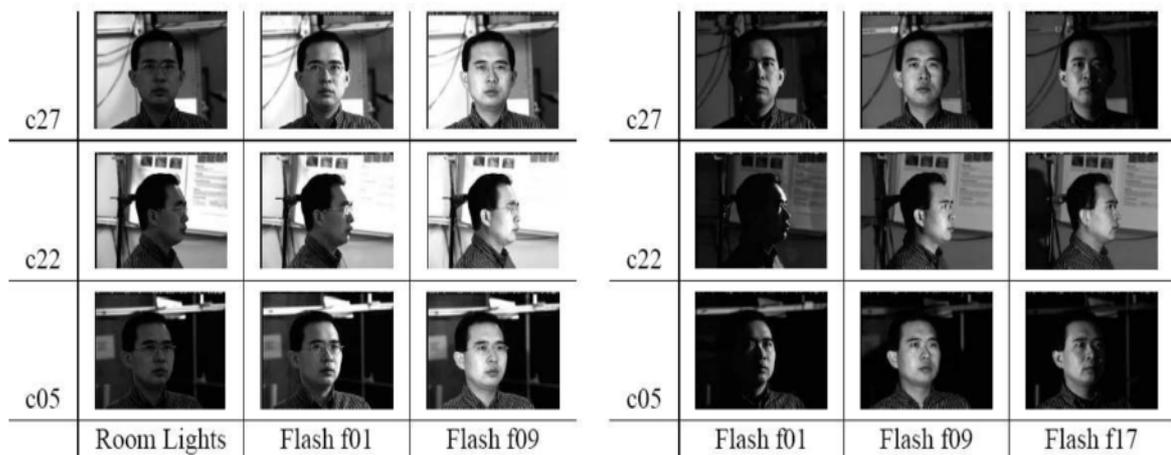


FIG47 : Extrait de la base PIE: à gauche des vues Illumination I et à droite des vues Illumination II

## 8. La base VISIO

La base VISIO [65] est une base de données contient actuellement des images nominale de 100 personnes. Les personnes sont répartis uniformément entre les sexes (Soit 50 personnes de sexe féminin et 50 personnes de sexe masculin). L'âge varie entre 19

et 69 ans. Les images sont prises dans des conditions contrôlées dans notre le laboratoire de Satya Wacana Christian University (SWCU). Les images sont prises sur un fond blanc uniforme en utilisant une seule caméra et des paramètres identiques. Une source de lumière fluorescente unique est placée en face de l'objet. L'appareil photo balance des blancs automatique est utilisée pour tous les personnes. Pour chaque personne, 105 images sont pris. Chaque photo a une combinaison différente du point de vue, l'expression du visage, et des accessoires. Ces variations sont résumées dans le tableau 8. Au total, la base de données VISIO Multiview contient 10500 images différentes. Chaque image dans la base de données manuellement Cultivées cropped autour de la zone du visage. Ensuite, les images sont rééchantillonnées dans un 64x64 pixels 8-bit image en niveaux de gris. La figure 46 montre les variations des images pour une personne.

Variations	
Point de vue	Niveau, bas, haut, frontal (0 °), 45 ° à droite, à droite 90 °, 45 ° à gauche, à gauche
Accessoires	Lunettes, Cap
L'expression du visage	Sourire, rire, neutre

**Tableau8** : Les variations dans la base de données VISIO.



FIG48: Variation des images pris pour une personne dans la base VISIO

## Annexe B

# Dualité lagrangienne et passage à la forme duale

### 1. Introduction

Cette annexe concerne la notion de dualité Lagrangienne. Elle montre la relation qui existe entre les deux formes primale et duale. Elle démontre comment le fait de résoudre le problème dual reviendrait à résoudre le problème primal. Nous verrons par la suite comment on fait le passage à la forme duale du système qui cherche à trouver l'hyperplan de séparation le plus optimal.

### 2. Dualité Lagrangienne

Soit un problème d'optimisation quadratique convexe qui s'écrit de la façon suivante :

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq 0 \quad \forall i = 1, \dots, n \end{cases}$$

Où  $f, C_1, C_2, \dots, C_n$  sont  $n + 1$  fonctions convexes définies sur un espace de Hilbert  $H$ .

Voici quelques notations qu'on utilisera au cours de cette section pour faciliter l'écriture des équations mathématiques :

si  $x \in \mathbb{R}^n$  alors  $x \geq 0 \Leftrightarrow x_i > 0 \quad \forall i \in 1, 2, \dots, n$

De la même manière si  $x \in \mathbb{R}^n$  alors  $x < 0 \Leftrightarrow x_i < 0 \quad \forall i \in 1, 2, \dots, n$

On peut supprimer parfois les crochets :  $\mathbf{ax}$  et  $\langle \mathbf{a}, \mathbf{x} \rangle$  représentent la même valeur mathématique.

Enfin si  $C$  est une fonction d'un espace de Hilbert  $H$ .

$$\{C \leq \mathbf{0}\} = \{x \in H \text{ tq } C(x) \leq \mathbf{0}\}$$

Le Lagrangien de la fonction  $f(x)$  noté  $L(x, a)$  est donné par la formule suivante :

$$L(x, a) = L(x, a_1, a_2, \dots, a_n) = f(x) \sum_i a_i c_i(x) = f(x) + aC(x)$$

Nous allons à présent énoncer quelques propositions qui vont nous permettre d'affirmer que résoudre le problème dual revient à résoudre le problème primal.

### Proposition 1 :

Soit  $x^* \in H$ ,  $a^* \geq \mathbf{0}$  et si  $\forall x \in H, a \geq \mathbf{0}$  alors  $L(x^*, a) \leq L(x^*, a^*) \leq L(x, a^*)$

Avec  $H$  un espace de Hilbert qui est une généralisation de l'espace euclidien.

Autrement dit :

$$f(x^*) + aC(x^*) \leq f(x^*) + a^*C(x^*) \leq f(x) + a^*C(x)$$

Alors  $x^*$  est une solution du problème :

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq \mathbf{0} \quad \forall i = 1, \dots, n \end{cases}$$

Et  $(x^*, a^*)$  est dit point selle du lagrangien.

La démonstration de cette proposition n'est pas très difficile.

On suppose  $M = \cap \{c_i \leq 0\}$ ,  $M \subset H$  (c'est l'ensemble des  $x$  qui vérifient toute les contraintes du problème) en séparant l'inégalité de gauche on a :

$$f(x^*) + aC(x^*) \leq f(x^*) + a^*C(x^*)$$

Donc  $aC(x^*) \leq a^*C(x^*)$  (I)

On a supposé que  $a^* \geq \mathbf{0}$  et comme  $x^*$  vérifie les contraintes du problème qu'on veut résoudre alors on déduit que  $C(x^*) \leq \mathbf{0}$ . En multiplions  $a^*$  par  $C(x^*)$  nous

obtenons :  $a^*C(x^*) \leq \mathbf{0}$  (II)

En résolvant de la même manière nous obtenons :

$$aC(x^*) \leq \mathbf{0} \quad \text{(III)}$$

De (I) et (II) et (III) on déduit que :

$$a^*C(x^*) = \mathbf{0}$$

Passons maintenant à l'inégalité de droite, on a :

$$f(x^*) + a^*C(x^*) \leq f(x) + a^*C(x)$$

Or

$$a^*C(x^*) = 0 \text{ donc } f(x^*) \leq f(x) + a^*C(x)$$

Nous pouvons prouver facilement comme on l'a fait précédemment que  $a^*C(x) \leq 0$ , ce résultat va nous permettre d'affirmer que :

$$f(x^*) \leq f(x) + a^*C(x)$$

On en déduit alors que :

$$f(x^*) \leq f(x) \quad \forall x \in M$$

Nous venons de démontré maintenant que  $x^*$  est la solution du problème

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq 0 \quad \forall i = 1, \dots, n \end{cases}$$

### Proposition 2:

$C_1, C_2, \dots, C_n: H \rightarrow R, n$  fonctions convexes et  $M$  est l'ensemble des  $x$  qui vérifient tous les contraintes du problème  $M = \cap \{C_i \leq 0\}$ . Telle que les deux contraintes équivalentes suivantes sont vérifiées :

$$\begin{aligned} & \exists x \in H \text{ tq } \forall i \in 1 \dots n, C_i(x) < 0 \\ & \forall a \geq 0, a \neq 0, \exists x \in M \text{ tq } \sum_i a_i C_i(x) < 0 \end{aligned}$$

Soit  $f$  une fonction convexe définie sur  $c$ . si  $x^*$  est une solution du problème :

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq 0 \quad \forall i = 1, \dots, n \end{cases}$$

Alors il existe  $a^* \in R^n, a \geq 0$  telle que :

$$\forall x \in H, a \geq 0, L(x^*, a) \leq L(x^*, a^*) \leq L(x, a^*)$$

La démonstration de cette proposition est plus difficile que celle de la première proposition. Les deux contraintes équivalentes sont appelées contraintes qualifiées, on pourra prouver leur équivalence en admettant le théorème suivant :

**Théorème :** Soit  $M$  un ensemble convexe ne contenant pas l'origine,

$$\exists a \in R^n, a \neq 0 \text{ telle que } \forall x \in M, 0 \leq \langle a, x \rangle$$

Grace à ce théorème on démontre que :

$$\begin{aligned} & \forall a \geq 0, a \neq 0, \exists x \in M \text{ tq } \langle a, x \rangle < 0 \text{ Alors} \\ & \exists x \in H \text{ tq } x < 0, \exists \text{ ensemble convexe } \in R^n \end{aligned}$$

Procédant par l'absurde pour prouver ce résultat.

Posons  $Q_x = \{y \in \mathbb{R}^2, y \geq 0 \text{ telle que } x < y\}$

Géométriquement l'ensemble  $Q_x$  est un quadrant ouvert en haut et à droite de  $x$

Soit  $Q = \bigcup_{x \in E} Q_x$   $Q$  est un ensemble convexe. S'il ne contient pas l'origine alors selon le théorème précédent,

$$\exists a \in \mathbb{R}^n, a \neq 0 \text{ telle que } 0 \leq \langle a, x \rangle \text{ pour tous } x \in E.$$

Ce qui est absurde avec notre hypothèse ( $\langle a, x \rangle < 0$ ). En on déduit donc que les contraintes qualifiées sont équivalentes.

Revenons maintenant à la démonstration de la proposition 2. Si  $x^*$  est une solution du problème :

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq 0 \quad \forall i = 1, \dots, n \end{cases}$$

Alors il n'existe aucun  $x \in H$  vérifiant les inégalités suivantes :

$$\begin{cases} f(x) - f(x^*) < 0 \\ C_1(x) \leq 0 \\ \vdots \\ C_n(x) \leq 0 \end{cases}$$

Cette affirmation vient du fait que  $\forall x \in H f(x) - f(x^*) \geq 0$ .

En on déduit que  $\forall x \in H f(x) - f(x^*) + \sum_{i=1}^n C_i(x) < 0$ .

Multiplions le terme précédent par  $\beta^* = (\beta_0^*, \dots, \beta_n^*) \in \mathbb{R}^{n+1}, \beta^* \geq 0$ .

Nous obtenons le résultat suivant :

$$\forall x \in H \beta_0^* (f(x) - f(x^*)) + \sum_{i=1}^n \beta_i^* C_i(x) \leq 0.$$

L'hypothèse d'équivalence des contraintes implique que  $\beta_0$  est non nulle, on pose  $a_i^* \beta_i^* / \beta_0$ , on obtient donc l'inégalité suivante :

$$\forall x \in H, f(x^*) \leq f(x) + \sum_{i=1}^n a_i^* C_i(x)$$

En posant  $x = x^*$  nous obtenons alors :

$$\sum_{i=1}^n a_i^* C_i(x^*) \geq 0 \quad (IV)$$

L'hypothèse d'équivalence des contraintes implique que

$$\forall x \in H, \forall a \in \mathbf{R}^n, a \geq 0, \sum_{i=1}^n a_i^* C_i(x^*) \leq 0 \quad (V)$$

De (IV) et (V) en on déduit que :

$$\sum_{i=1}^n a_i^* C_i(x^*) = 0 \quad (VI)$$

L'hypothèse d'équivalence des contraintes implique que :

$$aC(x^*) \leq 0$$

En on déduit donc des deux résultats précédents que :

$$aC(x^*) \leq a^*C(x^*)$$

En ajoutant  $f(x^*)$  à l'inégalité précédente nous obtenons :

$$f(x^*) + aC(x^*) \leq f(x^*) + a^*C(x^*)$$

D'où  $L(x^*, a) \leq L(x^*, a^*)$

Nous avons  $x^*$  solution du problème :

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq 0 \quad \forall i \in 1, \dots, n \end{cases}$$

Implique que :  $f(x^*) \leq f(x)$

L'hypothèse d'équivalence des contraintes implique que :

$$a^*C(x) \leq 0 \quad (VII)$$

De (VI) et (VII) en on déduit que :

$$a^*C(x) \leq a^*C(x^*)$$

Il en résulte que :

$$f(x^*) + a^*C(x^*) \leq f(x) + a^*C(x)$$

D'où  $L(x^*, a^*) \leq L(x, a^*)$

Nous sommes arrivé maintenant à démontrer que si  $x^*$  solution du problème :

$$\begin{cases} \min_{x \in H} f(x) \\ C_i(x) \leq 0 \quad \forall i = 1, \dots, n \end{cases}$$

Alors  $L(x^*, a) \leq L(x^*, a^*) \leq L(x, a^*)$

**Proposition 3**

On suppose que :

- $f, C_1, C_2, \dots, C_n$   $n + 1$  fonctions convexe de  $H$  dans .
- $f$  est strictement convexe.
- Pour chaque  $\mathbf{a} \geq \mathbf{0}$ ,  $L(x, \mathbf{a})$  est borné inférieurement.
- Le problème d'optimisation possède une solution  $x^*$ .

Nous posons :

$$- v(\mathbf{a}) = \min_{x \in H} L(x, \mathbf{a}) = \min_{x \in H} f(x) + \mathbf{a}C(x) \quad \text{et } M = \cap \{C_i \leq 0\}$$

Sous ces hypothèses :

- la solution  $x^*$  est unique.
- Il existe  $\mathbf{a}^* \geq \mathbf{0}$  tel que  $(x^*, \mathbf{a}^*)$  soit un point selle de  $L(x, \mathbf{a})$ .
- $\max_{\mathbf{a} \geq \mathbf{0}} (v(\mathbf{a})) = \min_{x \in c} (f(x))$

Les hypothèses de cette proposition ne sont pas redondantes. Ce n'est pas parce que  $f$  est bornée inférieurement que  $f$  atteint son minimum. De plus, enlever la condition que  $f$  est strictement convexe entraine que la solution n'est pas unique. Grace aux résultats de cette proposition, nous comprenons le lien qui est entre le problème primal et dual et comment le fait de résoudre le problème dual entraine la résolution du problème primal.

**Exemple :**

Afin de mieux comprendre les résultats énoncés précédemment voyons cet exemple. Soit à résoudre le système suivant :

$$\begin{cases} \min_{x,y \in \mathbb{R}} f(x,y) = x^2 + y^2 \\ 1 \leq x + y \end{cases}$$

En réécrivant ce system sous la forme standard vue précédemment nous obtenons le système suivant :

$$\begin{cases} \min_{x,y \in \mathbb{R}} f(x,y) = x^2 + y^2 \\ 1 - x - y \leq 0 \end{cases}$$

Le Lagrangien du système précédent est

$$L(x, y, \mathbf{a}) = x^2 + y^2 + \mathbf{a}(1 - x - y)$$

On cherche à calculer  $v(a)$  le maximum de  $L(x, y, a)$  par rapport à  $(x, y)$ . Une des méthodes de l'obtenir est d'utiliser les points critiques.

$$\frac{\partial L}{\partial x} = 2x - a = 0$$

d'où  $x = a/2$

$$\frac{\partial L}{\partial y} = 2y - a = 0$$

d'où  $y = a/2$

En remplaçant  $x, y$  par  $a/2$  dans le Lagrangien nous obtenons que :

$$v(a) = \frac{a^2}{4} + \frac{a^2}{4} + a(1 - a)$$

$$v(a) = \frac{a^2}{4} + a - a^2$$

$$v(a) = a(1 - a/2)$$

Pour résoudre ce problème maintenant, il suffit de trouver  $\max_a v(a)$  et de calculer par la suite  $x$  et  $y$ . Dans cet exemple il est simple de trouver  $\max_a v(a)$ . En passant par la dérivée première Nous aurons :

$$\frac{\partial v}{\partial a} = 1 - a = 0 \Rightarrow a = 1$$

En on déduit donc que  $a = 1$  est un optimum de  $v(a)$ . Il reste a déterminé si c'est un minimum ou un maximum.

On a que :  $v(2) = 2(1 - 1) = 0$

Donc  $v(2) < v(1)$

De tous cela il en résulte que  $a = 1$  est un maximum de  $v(a)$ .

Comme  $x = a/2$  et  $y = a/2$  alors  $(x, y) = \left(\frac{1}{2}, \frac{1}{2}\right)$  est la solution unique de notre problème.

### 3. Transformation du problème d'apprentissage en forme duale

#### 3.1 Cas linéairement séparables

Nous allons dans cette section appliquer la transformation précédente au système qu'on a obtenu en formulons les contraintes du cas linéairement séparables, mais faisons d'abord le rappel de ce système :

$$\begin{cases} \min_{\mathbf{w} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{w}\|^2 \\ \mathbf{y}_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b}) \geq 1 \quad \forall i \in 1 \dots n \end{cases}$$

Nous pouvons réécrire ce système en passant par la forme standard de la manière suivante:

$$\begin{cases} \min_{\mathbf{w} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{w}\|^2 \\ \mathbf{y}_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b}) \leq 0 \quad \forall i \in 1 \dots n \end{cases}$$

En passant par la forme lagrangienne, nous obtenons l'équation suivante :

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a}) = \frac{1}{2} \|\mathbf{w}\|^2 + \sum_i \mathbf{a}_i (1 - \mathbf{y}_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b}))$$

Nous cherchons à calculer  $\mathbf{v}(\mathbf{a})$  le maximum de  $L(\mathbf{w}, \mathbf{b}, \mathbf{a})$  par rapport à  $(\mathbf{w}, \mathbf{b})$ . Une des méthodes de l'obtenir est d'utiliser les points critiques.

Pour maximiser  $L$  par rapport à  $\mathbf{w}$  et  $\mathbf{b}$  on annule donc les dérivées partielles.

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i \mathbf{x}_i = \mathbf{0}$$

En en déduit alors que  $\mathbf{w} = \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i \mathbf{x}_i$

$$\frac{\partial L}{\partial \mathbf{b}} = \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i = \mathbf{0}$$

En remplaçant dans le lagrangien

$$\begin{aligned} L(\mathbf{w}, \mathbf{b}, \mathbf{a}) &= \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + \sum_i \mathbf{a}_i (1 - \mathbf{y}_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b})) \\ &= \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + \sum_{i=1}^n \mathbf{a}_i - \mathbf{a}_i \mathbf{y}_i \langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{a}_i \mathbf{y}_i \mathbf{b} \\ &= \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + \sum_{i=1}^n \mathbf{a}_i + \mathbf{b} \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i + \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i \langle \mathbf{w}, \mathbf{x}_i \rangle \end{aligned}$$

Comme  $\sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i = \mathbf{0}$  le paramètre  $\mathbf{b}$  va disparaître. De cette équation, en introduisant un nouveau paramètre  $\mathbf{j}$  pour pouvoir faire le produit vectorielle. Nous obtenons alors

$$\begin{aligned} &= \frac{1}{2} \sum_{i,j=1}^n \mathbf{y}_i \mathbf{y}_j \mathbf{a}_i \mathbf{a}_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i,j=1}^n \mathbf{y}_i \mathbf{y}_j \mathbf{a}_i \mathbf{a}_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^n \mathbf{a}_j \\ &= \sum_{i=1}^n \mathbf{a}_j - \frac{1}{2} \sum_{i,j=1}^n \mathbf{y}_i \mathbf{y}_j \mathbf{a}_i \mathbf{a}_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \end{aligned}$$

Nous pouvons maintenant formuler le problème dual :

$$\begin{cases} \max v(\mathbf{a}) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j a_i a_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \mathbf{a}_i \geq \mathbf{0} \text{ pour } i \in 1, \dots, n \end{cases}$$

La condition  $\mathbf{a}_i \geq \mathbf{0}$  pour  $i \in 1, \dots, n$  vient des suppositions que nous avons faites pour pouvoir passer à la forme duale. Le problème dual est plus simple à résoudre que le primal. La résolution de ce système consiste donc à calculer les  $\mathbf{a}_i$  à partir de ces derniers, on calcule  $\mathbf{w} = \sum_{i=1}^n \mathbf{a}_i y_i \mathbf{x}_i$ . Mais le terme  $\mathbf{b}$  n'apparaît pas en fonction des  $\mathbf{a}_i$ . Pour pouvoir calculer ce dernier nous devons utiliser les variables primales.

Nous admettons que :

$$\mathbf{b} = - \frac{\max_{y_i=-1}(\langle \mathbf{w}, \mathbf{x}_i \rangle) + \max_{y_i=1}(\langle \mathbf{w}, \mathbf{x}_i \rangle)}{2}$$

Il y'a une autre manière d'obtenir le  $\mathbf{b}$  mais on ne pourra utiliser cette méthode que dans le cas où les données sont linéairement séparables.

Si  $\mathbf{a}_i$  réalise l'optimum du problème dual,  $\mathbf{w}$  et  $\mathbf{b}$  celle du primal alors :

$$\sum_{i=1}^n \mathbf{a}_i (1 - y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b})) = \mathbf{0}$$

Cette équation vient du fait qu'on a prouvé que dans la deuxième et première proposition que :

$$\sum_{i=1}^n \mathbf{a}_i^* \mathbf{C}_i(\mathbf{x}^*) = \mathbf{0}$$

Comme Nous avons :

$$\sum_{i=1}^n \mathbf{a}_i (1 - y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b})) = \sum_{i=1}^n \mathbf{a}_i \mathbf{b} + \sum_{i=1}^n \mathbf{a}_i (1 - y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle)) = \mathbf{0}$$

D'où :

$$\mathbf{b} = - \frac{\sum_{i=1}^n \mathbf{a}_i (1 - y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle))}{\sum_{i=1}^n \mathbf{a}_i \mathbf{b}}$$

En ayant calculé les  $\mathbf{a}_i$  et  $\mathbf{b}$  nous pouvons à présent exprimer la fonction de décision de notre classificateur  $f(\mathbf{x}) = \sum_{i=1}^n \mathbf{a}_i y_i \langle \mathbf{x}, \mathbf{x}_i \rangle + \mathbf{b}$

### 3.2. Cas de données non linéairement séparables :

#### 3.2.1 Cas des fonctions noyaux :

Nous allons dans cette section appliquer la transformation précédente au système qu'on a obtenu en formulons les contraintes du cas linéairement séparables. Mais faisons d'abord le rappel de ce système :

$$\begin{cases} \min_{w \in \mathbb{R}^n} \frac{1}{2} \|w\|^2 \\ y_i (\langle w, \phi(x_i) \rangle + b) \geq 1 \quad \forall i \in 1 \dots n \end{cases}$$

En procédant de la même manière que lord du cas linéairement séparable, on obtient le problème dual suivant :

$$\begin{cases} \max v(a) = \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j k(\langle x_i, x_j \rangle) \\ a_i \geq 0 \end{cases}$$

Nous remarque que c'est le même problème que dans le cas linéaire. La seule différence c'est qu'on remplace  $\langle x_i, x_j \rangle$  par  $k(\langle x_i, x_j \rangle)$ . Ou on rappelle que  $k(\langle x_i, x_j \rangle) = \phi(x_i) * \phi(x_j)$ . Et qu'on a pour des raisons de l'immense temps de calcul du produit scalaire définit la fonction noyau qui nous permet de calculer ce produit sans connaître le mapping  $\phi$  qui l'a engendré. En remplaçant tous les occurrences du produit scalaire par le noyau, la fonction de décision  $f(x)$  serait :

$$f(x) = \sum_i a_i y_i k(x_i, x) + b$$

Telle que :

$$b = - \frac{\max_{y_i=-1} (\sum_i a_i y_i k(x_i, x)) + \max_{y_i=1} (\sum_i a_i y_i k(x_i, x))}{2}$$

ça reste toujours un problème de programmation quadratique simple à résoudre.

#### 3.2.2 Cas de l'introduction des variables de pénalités :

Dans le cas ou on a introduit les variables de pénalités. Ces dernières vont s'annuler en passant par la dérivée pour calculer  $v(a)$ . On en obtient donc le même problème que le problème dual, mais les conditions sur la résolution du système change. En effet, on va

suggérer que les  $\mathbf{a}_i$  qu'on va appeler les multiplicateurs de Lagrange soit non seulement positif ou nulle mais qu'il soit aussi inférieur au paramètre  $C$ . Voyons à présent comment on arrive à ces résultats. Rappelons que dans ce cas le problème primal est :

$$\left\{ \begin{array}{l} \min_{\mathbf{w}, \varepsilon} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum \varepsilon_i \\ \mathbf{y}_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b}) \geq 1 - \varepsilon_i \quad \forall i \in 1 \dots n \\ \varepsilon_i \geq 0 \quad \forall i \in 1 \dots n \end{array} \right.$$

On réécrit le problème de la manière standard

$$\left\{ \begin{array}{l} \min_{\mathbf{w}, \varepsilon} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum \varepsilon_i \\ 1 - \varepsilon_i - \mathbf{y}_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b}) \geq 0 \quad \forall i \in 1 \dots n \\ -\varepsilon_i \leq 0 \quad \forall i \in 1 \dots n \end{array} \right.$$

En passant par la forme duale nous obtenons :

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum \varepsilon_i + \sum_i \mathbf{a}_i (1 - \varepsilon_i - \mathbf{y}_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + \mathbf{b})) - \sum \mathbf{u}_i \varepsilon_i$$

Passons alors au point critique :

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i \mathbf{x}_i = \mathbf{0}$$

En on déduit alors que  $\mathbf{w} = \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i \mathbf{x}_i$

$$\frac{\partial L}{\partial \mathbf{b}} = \sum_{i=1}^n \mathbf{a}_i \mathbf{y}_i \mathbf{x}_i = \mathbf{0}$$

En dérivant sur chacune des  $\varepsilon_i$  nous obtenons :

$$\frac{\partial L}{\partial \varepsilon_i} = C - \mathbf{a}_i - \mathbf{u}_i = \mathbf{0}$$

Comme  $\mathbf{u}_i \geq \mathbf{0}$  alors  $C - \mathbf{a}_i \geq \mathbf{0}$

D'où vient la condition  $\mathbf{a}_i \leq C$

En remplaçant dans le Lagrangien est après plusieurs simplifications, nous arrivons à :

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a}) = \sum_{i=1}^n \mathbf{a}_i - \frac{1}{2} \sum_{i,j=1}^n \mathbf{y}_i \mathbf{y}_j \mathbf{a}_i \mathbf{a}_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^n (C - \mathbf{a}_i - \mathbf{u}_i) \varepsilon_i$$

Or comme  $C - \mathbf{a}_i - \mathbf{u}_i = \mathbf{0}$

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a}) = \sum_{i=1}^n \mathbf{a}_i - \frac{1}{2} \sum_{i,j=1}^n \mathbf{y}_i \mathbf{y}_j \mathbf{a}_i \mathbf{a}_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^n \mathbf{0} * \varepsilon_i$$

Donc

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a}) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j} y_i y_j a_i a_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

Nous obtenons le même lagrangien que dans le cas précédent, mais sous une nouvelle condition.

$$\begin{cases} \max v(\mathbf{a}) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j a_i a_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \mathbf{0} \leq a_i \leq C \text{ pour } i \in 1 \dots \dots \dots, n \end{cases}$$

Suivant les gammes de valeurs des multiplicateurs de Lagrange trois cas sont possibles :

- $a_i = 0$  alors l'exemple  $\mathbf{x}_i$  associé à cet  $a_i$  est bien classé et il ne se trouve pas dans un des hyperplans canoniques.
- $0 < a_i < C$  alors l'exemple  $\mathbf{x}_i$  associé à cet  $a_i$  est bien classé et il n se trouve dans un des hyperplans canoniques. Ce point est appelé support vecteur d'où vient le nom de la méthode (support vector machine).
- $a_i = C$  alors l'exemple  $\mathbf{x}_i$  associé à cet  $a_i$  est mal classé (on n'oublie pas qu'on connaît la classe des exemples du training-set) et il ne se trouve pas dans un des hyperplans canoniques. Malgré cela, il est considéré comme un support vecteur.

# Bibliographie

- [1]: S. Liu, M. Silverman, " A Practical Guide to Biometric Security Technology", IEEE Computer Society, IT Pro-Security, Janvier-Février 2001.
- [2]: CHIUNHSIUN L., KUO-CHIN F. : Human Face Detection using Geometric triangle relationship, 15th International Conference on Pattern Recognition ICPR'2000, Barcelona, September 3-8, 2000.
- [3]: KAMEL M. S., SHEN H. C., WONG A. K. C. and CAMPEANU R. I.: System for the recognition of human faces, IBM Systems Journal, Vol. 32, No 2, p. 307-320, 1993.
- [4]: VEZJAK M., KOROSEC J., GYERGYEK L., PAVESIC N., SAVSEK T., ERJAVC I. and GERE A.: System for description and identification of individuals, IEEE, Vol. 2, p. 1251- 1254, 1991.
- [5]: VOGEL S., SCHMID C.: Face Detection on generic local descriptors and spatial constraints, Proceedings of the 15th International Conference on Pattern Recognition ICPR'2000, Vol. 1, p. 1084-1087, Barcelona, September 3-8, 2000.
- [6]: John D. Woodward Jr., Christopher Horn, Julius Gatune, and Aryn Thomas, "Biometrics A Look at Facial Recognition", documented briefing by RAND public Safety and Justice for the Virginia State Crime Commission, 2003.
- [7]: SOUHILA GUERFI ABABSA "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D", UNIVERSITE D'EVRY VAL D'ESSONNE 2008.
- [8]: L. Hong, A Jain, " Integrating Face and Fingerprints for Personal Identification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, n°12, p. 1295-1307, 1998.
- [9]: L. Hong, A Jain, S. Pankanti, " Can Multibiometrics Improve Performance?", Proceedings AutoID'99, Summit, NJ, p.59-64, Oct 1999.

- [10]: P.J. Phillips, H. Hyeonjoon, S. Rizvi, P. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vo. 22, No. 10, Octobre 2000.
- [11]: Kamel (M. S.), Shen (H. C.), Wong (A. K. C.) and Campeanu (R. I.) 1993. System for the recognition of human faces. *IBM systems Journal*, Vol. 32, No 2, page 307-320.
- [12]: INSPASS, <http://www.ins.usdoj.gov/praphics/howdoi/inspass.htm>.
- [13]: J. Bellegarda, D. Naik, M. Neeracher, K. Silverman, "Language-Independent, Short Enrollement Voice Verification over a Far-Field Microphone", *ICASSP*, Vol. 1, p. 445-448, Salt Lake City, Utah, 7-11 Mai 2001.
- [14]: " Biometrics Comes To Life", *Banking Journal*, Janvier 1997  
[http://www.banking.com/aba.cover\\\_0197.html](http://www.banking.com/aba.cover\_0197.html)
- [15]: T-NETIX Inc., <http://www.t-netix.com>.
- [16]: Siemens Automotive, <http://media.siemensauto.com>.
- [17]: P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* 16, 295–306, 1998b.
- [18]: S. A. Rizvi, P. J. Phillips, H.A. Moon. Verification protocol and statistical performance analysis for face recognition algorithms. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*. 833–838, 1998.
- [19]: K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre. XM2VTSDB: The Extended M2VTS Database. In *Proceedings, International Conference on Audio- and Video-Based Person Authentication*. 72–77, 1999.
- [20]: D. Blackburn, M. Bone, P. J Phillips. "Face recognition vendor test 2000". Tech. rep. <http://www.frvt.org> , 2001
- [21]: P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. Face recognition vendor test 2002: Evaluation report. NISTIR 6965. Available online at <http://www.frvt.org>, 2003.
- [22]: S. Prabhakar, J. Kittler, D. Maltoni, L. O’Gorman, and T. Tan. Introduction to the Special Issue on Biometrics: Progress and Directions , *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, April 2007.

- [23]: Zhao,W., Chellappa,R., and Krishnaswamy,A. Discriminant Analysis of Principal Components for Face Recognition. In Proceedings, International Conference on Automatic Face and Gesture Recognition. 336-341, 1998.
- [24]: Leroy,B., Chouakria,A., Herlin,L., et Diday,E. Approche Géométrique et Classification pour la Reconnaissance de Visage. INRIA, France, 1996
- [25]: M.T. Laskri, Dj Chefrour. Who\_ is : système d'identification des visages humains, Arima Volume 1-2002.
- [26]: Y. Adini, Y. Moses, S. Ullman, Face recognition: The problem of compensating for changes in illumination direction. IEEE Trans. Patt. Anal. Mach. Intell. 19, 721–732, 1997.
- [27]: D. Blackburn, M. Bone, P. J Phillips. "Face recognition vendor test 2000". Tech. rep. <http://www.frvt.org>, 2001
- [28]: Y.Moses, Y. ADINI et S. ULIMAN. "Face Recognition: the problem of Compensating Changes in Illumination Direction". Dans Proceedings of the European Conference on Computer Vision (ECCV), volume 1, pages 286-296, 1994.
- [29]: P. J. Phillips, P. J.Grother, R. J.Micheals, D. M. Blackbrun, E. Tabassi, and J. M. Bone. Face recognition vendor test 2002: Evaluation report. NISTIR 6965. Available online at <http://www.frvt.org>, 2003.
- [30]: R. Gross, J. Shi, J. Cohn. Quo Vadis Face Recognition?. Third Workshop on Empirical Evaluation Methods in Computer Vision, December, 2001.
- [31]: A.M Martínez, R. Banavente. The AR face database. Tech. Report 24 CVC Barcelone, Espagne, June 1998.
- [32]: P.J. PHILLIPS, P. GROTHER, R.,J MICHEALS? D.M. BLACKBURN, E. TABASSI et J.M. BONE. " Face Recognition Vendor Test 2002. Evaluation Report.". Technical Report 6965, National Institute of Standards and Technology, 2003. 56 pages.

- [33]: Pentland,A., Moghaddam,B., and Starner,T. View Based and Modular Eigenspaces for Face recognition. IEEE Conference on Computer Vision and Pattern Recognition, 1994.
- [34]: Harmon,L.D.,Khan,M.K.,Lasch,R,and Raming,P.F. Machine Identification of Human Faces. Pattern Recognition, 13(2):97-110, 1981.
- [35]: Brunelli,R., and Poggio,T. Face Recognition:Features versus Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.15, pp. 1042-1052,1993.
- [36]: Leroy,B., Chouakria,A., Herlin,L., et Diday,E. Approche Géométrique et Classification pour la Reconnaissance de Visage. INRIA, France, 1996
- [37]: Cox,I.J., Ghosn,J., and Yianilos,P.N. Feature-Based Recognition Using Mixture-Distance. In Proceedings,IEEE Conference on Computer Vision and pattern Recognition. 209-216, 1996.
- [38]: Kamel M.S, Shen H.C, Wong A.K.C, Campeanu R.I:system for the recognition of human faces, IBM Systems Journal, Vol.32, No 2, P. 307-320,1993.
- [39]: W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face Recognition: A Literature Survey", ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 399-458.
- [40]: Xiaoyin xu, "Image Based Face Recognition using Global Features", Research Centre for Integrated Microsystems Electrical and Computer Engineering University of Windsor, Supervisors: Dr. Ahmadi Seminar, May 13, 2005. Document available at: <http://www.visi.uwindsor.ca:RCIM3:Seminar.html>
- [41]: V. BRUCE, P.J.B. HANCOCK et A.M. BURTON. "Humain Face Perception and Identification". Dans H. WECHSLER, P.J. PHILIPS, V. BRUCE, F.F. SOULIE et T.S. HUANG, éditeurs, Face Recognition: from Theory to Applications pages 51-72. Springer-Verlag, Berlin, Allemagne, 1998.

- [42]: A. Lemieux « Systèmes d'identification de personnes par vision numérique » Université Laval, Québec Décembre 2003.
- [43]: Andrew Webb, Statistical Pattern Recognition, Second Edition, 2002.
- [44]: David C. Lay, Linear Algebra And Its Applications, Second Edition, 2000.
- [45]: Richard Lepage : Reconnaissance d'algue toxique par vision artificielle et réseau de 04neurones, mémoire de recherche, université du québec à Rimouski, 2004.
- [46]: Ben Kröse, Patrick van der Smagt, An introduction to Neural Networks, Eighth Edition, November 1996
- [47]: "Support vector machine"  
Document available at: [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine).
- [48]: P. Sinha, "Object Recognition via Image Invariants: A Case Study," Investigative Ophthalmology and Visual Science, vol. 35, no. 4, pp. 1735-1740, 1994.
- [49]: P. Sinha, "Processing and Recognizing 3D Forms," PhD thesis, Massachusetts Inst. of Technology, 1995.
- [50]: Ferdinando Samaria: Face Recognition Using Hidden Markov Models. Thèse de doctorat, Engineering Department, Cambridge University, Trumpington Street, Cambridge CB2 1PZ, UK, October 1994.
- [51]: Ara V. Nefian et Monson H. Hayes III : An embedded hmm based approach for face detection and recognition. Dans IEEE International Conference on Acoustics, Speech and Signal Processing, volume VI, pages 3553–3556, March 1999.
- [52]: Alexandre Lemieux « Système d'identification de personnes par vision numérique », 2003, page 153
- [53]: KAMEL M. S., SHEN H. C., WONG A. K. C. and CAMPEANU R. I.: System for the recognition of human faces, IBM Systems Journal, Vol. 32, No 2, p. 307-320, 1993.

- [54]: M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, Malsburg von der, R. Wurtz. Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.* 42 (3) (1993) 300–311.
- [55]: J. Zhang, Y. Yan, and M. Lades. Face recognition : Eigenface, elastic matching, and neural nets. *PIEEE*, 85(9) :1423 1435, September 1997.
- [56]: Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :775 779, 1997.
- [57]: V. Vapnik. *The nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.
- [58]: C.J.C Burges. A tutorial on Support Vector Machines for patterns recognition. *Data Mining and Knowledge Discovery*, 2(2):121-167, 1998.
- [59]: A.J. Smola and B. Schölkopf. A tutorial on Support Vector Regression, *euroCOLT2 Technical Report Series*, NC2-TR-1998-030. October 1998
- [60]: N.D. Ayat, “Sélection Automatique de Modèle Dans Les Machines à Vecteurs de support: Application à La Reconnaissance d'Image De Chiffres Manuscrits“, PhD thesis, Université of Québec, 2004
- [61]: C.-W. Hsu and C.-J. Lin. A comparison of multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415-425, 2002.
- [62]: L. Bottou, C. Cortes, J. Denker, H.s Drucker, I. Guyon, L. Jackel, Y. Le Cun, U. Muller, E. Sackinger, P. Simard, and V. Vapnik. Comparison of classifier methods: A case study in handwritten digit recognition. In *Proceedings of 13th International Conference on Pattern Recognition*, 1994.
- [63]: J. Friedman. Another approach to polychotomous classification. Technical report, Departement of statistics, Stanford University, 1996.
- [64]: Shivani Gupta, K.S.Patnaik. Enhancing performance of face recognition System by using near set approach for Selecting facial features. *Journal of Theoretical and Applied Information Technology*. 2005 - 2008 JATIT.

- [65]: Linlin SHEN. Gabor Features and Support Vector Machine for Face Identification. *Biomedical Soft Computing and Human Sciences*, Vol.14, No.1, pp.61-66 (2009).
- [66]: Jan Mazanec, Martin Melisek, Milos Oravec, Jarmila Pavlovicova. Support Vector Machines, PCA And LDA in face recognition. *Journal of ELECTRICAL ENGINEERING*, VOL. 59, NO. 4, 2008, 203–209.
- [67]: Bing Luo, Yun Zhang, Yun-Hong Pan. Face Recognition Based on Wavelet Transform and SVM. *Proceedings of the 2005 IEEE International Conference on Information Acquisition June 27 - July 3, 2005, Hong Kong and Macau, China*
- [68]: Ivanna K. Timoutius, Iwan Setyawan, and Andreas A. Febrianto. Face Recognition between Two Person using Kernel Principal Component Analysis and Support Vector Machines. *International Journal on Electrical Engineering and Informatics – Volume 2, Number 1, 2010.*
- [69]: Zhifeng Li and Xiaou Tang. Bayesian Face Recognition Using Support Vector Machine and Face Clustering. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04).*
- [70]: Masoud Mazloom, Shohreh Kasaei, Hoda Alemi. Construction and Application of SVM Model and Wavelet-PCA for Face recognition. *2009 Second International Conference on Computer and Electrical Engineering.*
- [71]: Ergun Gumus, Niyazi Kilic, Ahmet Sertbas, Osman N. Ucan. Eigenfaces and Support Vector Machine Approaches for Hybrid Face Recognition. *The Online Journal on Electronics and Electrical Engineering (OJEEE) Vol. (2) – No. (4) Reference Number: W10-0005.*
- [72]: Hervé BREDIN, Najim DEHAK, Gérard CHOLLET. GMM-based SVM for face recognition. 0-7695-2521-0/06/\$20.00 (c) 2006 IEEE.
- [73]: D.G. Lowe. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol.60, pp.91–110, 2004.

- [74]: T.Lindeberg, Feature detection with automatic scale selection, *International Journal of Computer Vision*, vol.30, no. 2, pp. 79–116, 1998.
- [75]: R.GROSS. "Face Databases". Dans S.Z. LI et A.K. JAIN, éditeurs *Handbook of Face Recognition*, Chapitre 13. Springer-Verlag, Reidel, Dordrecht, 2004. 22 Pages.
- [76]: P.J. PHILLIPS, H.MOON, S.A. RIZVI et P.J. RAUSS. " The FERET Evaluation Methodology for Face- Recognition Algorithms". *IEEE Transactions on Pattern Analysis and machine Intelligence*, 22:1090 1104, 2000.
- [77]: P.N. BELHUMEUR, J.P. HESPANHA et D.J. KRIEGMAN. "Eigenface vs Fisher-face: Recognition Using Class Specific Linear Projection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711-720, 1997.
- [78]: F.SAMARIA et HARTER, "Parametresation of a Stochastic Model for Human Face Identification". Dans *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 138-142, 1994.
- [79]: B.W. HWANG, M.C. ROH et S.W. LEE. " Performance Evaluation of face Recognition Algorithms on Asian Face Database". Dans *Proceedings of the IEEE international Conference on automatic Face and Gesture Recognition (FGR)*, pages 278 283, 2004.
- [80]: D.GRAHAM et N. ALLISON. " Characterizing Virtual Eigensignatures for General Purpose Face Recognition". Dans H.WECHSLER, P.J. PHILLIPS, V. BRUCE, F.F. SOULIE et T.S. HUNG, éditeurs, *Face Recognition: from Theory to Applications*, pages 446-456. Springer-Verlag, Berlin, Allemagne, 1998.
- [81]: A.R.MARTINEZ et R.BENAVENTE. " The AR Database". Technical Report 24, Computer Vision Center (CVC), Barcelone, Espagne, 1998. [82]: T. SIM, S. BAKER et M. BSAT. " The CMU Pose, Illumination, and Expression Database". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615-1618, 2003.