



# Avant-propos

*Le présent document constitue le mémoire de la thèse « Détection et Reconnaissance de visage dans un document mixte » dans le contexte de l'obtention du Doctorat de l'Université Sultan Moulay Slimane - Beni Mellal - spécialité Informatique Décisionnelle.*

# Remerciements

Je tiens, d'abord, à remercier Mr M. Oukessou, de m'avoir permis d'intégrer le domaine de la recherche. Je le remercie pour son aide précieuse et son encouragement.

Je voudrai témoigner mes sincères remerciements et ma profonde reconnaissance à Mr R. El Ayachi pour m'avoir encadré durant la durée de réalisation de ce travail. Il a été toujours présent pour écouter mes propositions et répondre à mes questions. Il m'a orienté constamment vers la bonne direction. Je le remercie beaucoup pour la qualité de son encadrement, son soutien moral, ses conseils et ses qualités humaines.

Merci également à Mr M. Baslam, pour sa confiance, sa disponibilité et son soutien continu. Qu'il trouve ici l'expression de ma gratitude pour ses précieux conseils et toute l'aide qu'il m'a procurée durant l'élaboration de ce travail. Je tiens à le remercier bien vivement pour son encadrement, son encouragement et aussi pour sa contribution à l'amélioration judicieuse de la qualité de ce mémoire.

Je remercie vivement les membres de jury d'avoir accepté de juger mon travail de thèse. Je suis très reconnaissant aux rapporteurs de consacrer du temps pour se pencher avec rigueur et grand intérêt sur ce rapport et de m'avoir fait l'honneur d'en être les rapporteurs. Mes plus sincères remerciements vont aussi à l'examineur qui a examiné ce travail de thèse et a participé au jury.

Merci en particulier à Mr et Mme Mlliani, pour leurs conseils, leurs encouragements et leurs directives le long de cette thèse. Je remercie également l'ensemble du personnel du département de l'informatique leurs accueils chaleureux et les conseils pédagogiques dont ils m'ont fait part.

Merci également Mr M. Fakir pour m'avoir proposé le sujet de recherche, pour l'aide et l'encadrement qu'il m'a apporté avant et durant cette thèse.

Merci à tous mes amis et à mes proches pour leur soutien et leurs encouragements, plus particulièrement, ma mère Zahia, ma femme Fatiha mon fils Ali et ma princesse Rihabe, je les remercie tous du profond de mon cœur pour leur affection et leurs encouragements et pour m'avoir supporté pendant ces quelques années. Merci donc pour ces années de compréhension, de patience et d'efforts communs.

Je ne peux terminer mes remerciements sans exprimer ma sincère gratitude à tous mes collègues au Lydex, en particulier Mr C. Boeckel, Directeur des CPGE et à tous mes amis d'ECOSIG en particulier Mr A. Moustahfid et à ceux de Day-Prépas pour leurs encouragements et l'aide qu'ils m'ont procurée durant l'élaboration de cette thèse.

# Résumé

Nous présentons dans cette thèse un système d'identification de visage dans un document composé de textes et d'images (y compris des visages). Ce système est fondé, sur l'utilisation de trois méthodes : la méthode « Hybride », la méthode « Template Matching » et une méthode basée sur l'utilisation des propriétés géométriques. Le système se décompose de quatre phases successives de traitement : a)- Localisation des images, ayant la couleur de la peau b)- Localisation des visages dans ces images et c)- Evaluation du vecteur caractéristique des principaux éléments de chaque visage et l'analyse factorielle des données ainsi extraites.

La localisation des images contenant, probablement, des visages, dans le document est effectuée à l'aide d'une technique basée sur la couleur de la peau (Méthode Hybride).

Afin d'extraire ensuite des régions contenant les éléments significatifs du visage, tels que la bouche, les yeux et le nez, nous employons la méthode de Template Matching. Cette méthode est fondée sur l'utilisation d'un modèle pour chaque région, ce qui permet d'adapter simplement le modèle à la morphologie de chacun des objets du visage.

Les régions, ainsi obtenus du visage à identifier, sont ensuite mis en correspondance avec ceux obtenus sur les visages de la base d'apprentissage. A cet effet, nous proposons une nouvelle méthode géométrique étendue de la méthode WHO\_IS. Cette méthode a pour but de déterminer des données symboliques correspondant au vecteur caractéristique du visage.

Les résultats positifs des essais expérimentaux démontrent l'efficacité des algorithmes proposés pour construire un système de reconnaissance de visage plus performant.

**Mots clés :** *Méthode Template Matching ; Méthode Hybride ; Méthode Géométrique ; Détection et Reconnaissance de visage.*

# Abstract

In this paper, we present a face identification system in a document containing both text and graphics (faces included). This system is based on the use of three methods: the "Hybrid" method, the "Template Matching" method and a method based on the use of geometric properties. The system involves four successive phases of treatment: a) - Location of images having skin color b) - Location of the faces in such images and c) - Evaluation of the feature vector of the main elements of each face and the factor analysis of the data thus extracted.

The location of images, likely to contain faces in the document, is carried out using a technique based on skin color (Hybrid Method).

In order to then extract regions containing significant face features, such as the mouth, eyes and nose, we use the Template Matching method. The latter is based on the use of a model for each region, which enables us to simply adapt the model to the morphology of each part of the face.

The regions thus obtained from the face to be identified are then matched with those obtained on the faces of the training set. For this purpose, a new extended geometric method of the WHO\_IS type is proposed. This method aims to determine symbolic data corresponding to the feature vector of the face.

The positive results of the experimental tests demonstrate the efficiency of the proposed algorithms in order to build a more advanced face recognition system.

**Keywords:** Template Matching Method; Hybrid method; Geometric method; Face Detection and Recognition.

# Liste des publications et communications

## 1- Publications

### ☞ 1er publication:

Lhoussaine Bouhou<sup>1</sup>, Rachid El Ayachi<sup>2</sup>, Mohamed Fakir<sup>3</sup>, Mohamed Oukessou<sup>4</sup>, Recognition of a Face in a Mixed Document, TELKOMNIKA Indonesian Journal of Electrical Engineering, Vol. 15, No. 2, August 2015, pp. 301 ~ 312, DOI: 10.11591/telkomnika.v15i2.8373.

<sup>1,2,3</sup>Computer Sciences Departement, Faculty of Science and Technology, Sultan Moulay Slimane University, Morocco.

<sup>4</sup>Mathematics Department, Faculty of Science and Technology, Sultan Moulay Slimane University, Morocco.

<http://www.iaesjournal.com/online/index.php/TELKOMNIKA/article/view/8373/4093>

### ☞ 2ème publication:

Lhoussaine Bouhou<sup>1</sup>, Rachid El Ayachi<sup>2</sup>, Mohamed BASLAM<sup>3</sup>, Mohamed Oukessou<sup>4</sup>, Face detection in a mixed-subject document, [IJECE] #12725.

International Journal of Electrical and Computer Engineering (IJECE), Vol. 6, No. 6, December 2016, pp. 2828~2835, ISSN: 2088-8708, DOI: 10.11591/ijece.v6i6.12725

<sup>1,2,3</sup>Computer Sciences Departement, Faculty of Science and Technology, Sultan Moulay Slimane University, Morocco.

<sup>4</sup>Mathematics Departement, Faculty of Science and Technology, Sultan Moulay Slimane University, Morocco.

<http://iaesjournal.com/online/index.php/IJECE/article/view/12725>

### ☞ 3ème publication:

Lhoussaine Bouhou<sup>1</sup>, Rachid El Ayachi<sup>2</sup>, Mohamed Baslam<sup>3</sup> and Mohamed Oukessou<sup>4</sup>, Face Recognition in a Mixed Document based on the Geometric Method.

International Journal of Advanced Science and Technology, Vol.116 (2018), pp.109-120.

<sup>1,2,3</sup>Sultan Moulay Slimane University, Faculty of Science and Technology, Informatique Department.

<sup>4</sup>Sultan Moulay Slimane University, Faculty of Science and Technology, Mathematique Department.

<http://dx.doi.org/10.14257/ijast.2018.116.10>

## 2-Communications

↪ April 23-25, CBI'15 (The second International Conference on Business Intelligence)  
Exposé Oral le 24-04-2015 intitulé: « Reconnaissance de visage dans un document mixte ».

↪ Certificate of attendance - certify that Lhoussaine BOUHOU attended International Conference on Intuitionistic Fuzzy Sets Theory and Applications 20 – 22 April 2016.

↪ March 29-31, 2017, CBI'17 (The second International Conference on Business Intelligence).  
Exposé Oral le 31 March 2017 intitulé: « Détection et Reconnaissance de visage dans un document mixte ».

↪ April 25-27, 2018, CBI'18 (International Conference on Business Intelligence ). Exposé intitulé:  
«A Geometric method for facial recognition in a joint paper».

# Table des matières

<b>LISTE DES FIGURES.....</b>	<b>11</b>
<b>LISTE DES TABLEAUX.....</b>	<b>12</b>
<b>INTRODUCTION.....</b>	<b>13</b>
<b>MOTIVATION ET PROBLEMATIQUE.....</b>	<b>14</b>
<b>ARCHITECTURE DU SYSTEME PROPOSE.....</b>	<b>16</b>
<b>CONTRIBUTIONS ET PLAN DU MANUSCRIT.....</b>	<b>17</b>
<b>1. GENERALITES SUR LA RECONNAISSANCE DE VISAGE.....</b>	<b>20</b>
<b>1.1. INTRODUCTION.....</b>	<b>20</b>
<b>1.2. LE CONTROLE D'ACCES PAR BIOMETRIE.....</b>	<b>20</b>
1.2.1. TACHES DE L'ANALYSE DU VISAGE.....	21
1.2.2. LES COMPOSANTES D'UN SYSTEME BIOMETRIQUE.....	23
<b>1.3. EVALUATION DES PERFORMANCES D'UN SYSTEME.....</b>	<b>24</b>
<b>1.4. BASES DES DONNEES DISPONIBLES.....</b>	<b>25</b>
<b>1.5. DETECTION DE VISAGES : ETAT DE L'ART.....</b>	<b>26</b>
1.5.1. CLASSIFICATION DES METHODES DE LOCALISATION FACIALE.....	27
1.5.1.1. Approche globale sans segmentation.....	27
1.5.1.2. Approche avec segmentation.....	28
1.5.2. METHODE DE DETECTION DE VISAGE DE VIOLA & JONES.....	29
1.5.2.1. Éléments de la méthode.....	30
1.5.2.2. Caractéristiques pseudo-Haar.....	31
1.5.2.3. Approche d'image intégrale.....	31
1.5.2.4. Algorithme AdaBoost.....	31
1.5.2.5. Algorithme en cascades de classifieurs.....	31
1.5.2.6. Étapes clés de la méthode.....	32
1.5.3. METHODE DE DETECTION DE VISAGE BASEE SUR L'ANALYSE DE LA COULEUR DE LA PEAU.....	33
1.5.3.1. Les méthodes explicites.....	35
1.5.3.2. Les approches non paramétriques.....	36
1.5.4. TEMPLATE MATCHING.....	36
<b>1.6. RECONNAISSANCE FACIALE : ETAT DE L'ART.....</b>	<b>38</b>
1.6.1. PHASE D'EXTRACTION DU VECTEUR CARACTERISTIQUE DU VISAGE.....	40
1.6.1.1. La méthode d'Analyse en Composantes Principales(ACP).....	40
1.6.1.2. Moments de Hu.....	42
1.6.1.3. Moments de Legendre.....	44



1.6.1.4. Les Méthodes Géométriques .....	45
1.6.2. PHASE DE CLASSIFICATION .....	47
1.6.2.1. Types d'apprentissage .....	48
1.6.2.2. Choix de la Base d'apprentissage.....	49
1.6.2.3. Technique des réseaux de neurones .....	50
<b>CONCLUSION.....</b>	<b>52</b>

## **2. CONTRIBUTION A LA DETECTION DE VISAGE DANS UN DOCUMENT MIXTE ..... 54**

### **2.1. INTRODUCTION..... 54**

<b>2.2. APPROCHE DES HISTOGRAMMES .....</b>	<b>55</b>
2.2.1. NOTION D'HISTOGRAMME.....	55
2.2.2. DESCRIPTION DE L'APPROCHE.....	56
<b>2.3. APPROCHE BASEE SUR LES METHODES 'HYBRIDE' ET 'TEMPLATE MATCHING'.....</b>	<b>60</b>
<b>2.4. OBJECTIF DE LA DEUXIEME APPROCHE.....</b>	<b>61</b>
2.4.1. DETECTION DES SECTIONS AYANT LA COULEUR DE LA PEAU.....	62
2.4.2. : FILTRAGE DES REGIONS AYANT LA COULEUR DE LA PEAU.....	63
2.4.3. EXTRACTION DES REGIONS DES IMAGES COULEUR PEAU ET DETECTION DE VISAGE.....	64
<b>2.5. RESULTATS ET DISCUSSIONS .....</b>	<b>66</b>
2.5.1. INTERFACE GRAPHIQUE DU SYSTEME DEVELOPPE DANS SA PHASE DE DETECTION.....	66
2.5.2. RESULTATS DE L'APPROCHE DES HISTOGRAMMES.....	67
2.5.3. RESULTATS DE L'APPROCHE BASEE SUR LA METHODE HYBRIDE ET LA METHODE TEMPLATE .....	68
<b>CONCLUSION.....</b>	<b>69</b>

## **3. CONTRIBUTION A LA RECONNAISSANCE DE VISAGE DANS UN DOCUMENT MIXTE..... 71**

<b>3.1. INTRODUCTION .....</b>	<b>71</b>
<b>3.2. RECONNAISSANCE DE VISAGE : PREMIERE APPROCHE PROPOSEE ....</b>	<b>72</b>
3.2.1. EXTRACTION DES CARACTERISTIQUES DU VISAGE.....	73
3.2.2. CLASSIFICATION .....	73
3.2.2.1. Réseau de neurones .....	73
3.2.2.2. Les K-Plus proche voisin.....	74
3.2.3. RESULTATS ET DISCUSSION .....	74

3.2.4. ILLUSTRATION ET INTERPRETATION DES RESULTATS.....	77
<b>3.3. RECONNAISSANCE DE VISAGE :DEUXIEME APPROCHE PROPOSEE .....</b>	<b>77</b>
3.3.1. EXTRACTION DES CARACTERISTIQUES .....	77
3.3.2. CLASSIFICATION .....	80
APPROCHE D'EXTRACTION.....	81
COUCHE D'ENTREE .....	81
COUCHE CACHEE .....	81
COUCHE DE SORTIE .....	81
METHODE GEOMETRIQUE .....	81
17 NEURONES.....	81
10 NEURONES.....	81
40 NEURONES.....	81
<i>TABLEAU 3.2. CARACTERISTIQUES DU RESEAU DE NEURONES UTILISE PAR NOTRE</i>	
<i>SYSTEME.</i> .....	81
3.3.3. BASE DE DONNEES .....	82
<b>3.4. RESULTATS ET DISCUSSIONS .....</b>	<b>82</b>
<b>CONCLUSION.....</b>	<b>83</b>
 <b>CONCLUSION GENERALE ET PERSPECTIVES .....</b>	 <b>84</b>
 <b>REFERENCES .....</b>	 <b>86</b>

## Liste des figures

Figure 1 : Architecture de système de reconnaissance de visage dans un document mixte. ....	17
Figure 1.1 : Modèles de visage basé sur maillage de points pour la reconnaissance d'identité.....	28
Figure 1.2 : Sélection par un seul classifieur « fort ».....	30
Figure 1.3 : Sélection par boosting .....	32
Figure 1.4 : Schéma illustratif de la méthode Template Matching .....	37
Figure 1.5 : Extraction du contour des yeux / bouche.....	45
Figure 1.6 : Visage accompagné de mesures géométriques .....	46
Figure 1.7 : Modèle géométrique du visage dans le système WHO_IS .....	47
Figure 1.8 : Exemple simple de classification par un algorithme K-Plus Proches Voisins.....	49
Figure 1.9 : Analogie entre le neurone biologique et le neurone formel.....	51
Figure 2.1 : Synoptique de notre système dans sa phase de détection de visage .....	54
Figure 2.2 : Image Lena en codage RVB .....	55
Figure 2.3 : Composante R(a), composante V(b), composante B(c).....	55
Figure 2.4 : Histogramme R (a), histogramme V (b), histogramme B(c) .....	55
Figure 2.5 : Première vision décrivant la phase détection du système envisagé à développer.....	56
Figure 2.6 : Différents exemples de documents, utilisés, contenant un visage .....	57
Figure 2.7 : Extraction de l'image.....	58
Figure 2.8 : Détection du visage.....	59
Figure 2.9 : Description de la phase de détection du système dans la deuxième approche.....	60
Figure 2.10 : Processus reflétant l'objectif général du système développé.....	62
Figure 2.11 : Exemple de document mixte convertit en binaire développé par notre système .....	63
Figure 2.12 : Exemple de document filtré.....	64
Figure 2.13 : Exemple de localisation de deux images couleur peau dans un document mixte.....	64
Figure 2.14 : Modèle de Template utilisée pour détecter si une image contient un visage .....	65
Figure 2.15 : Détection de visage par méthode Template Matching.....	65
Figure 2.16 : Segmentation de l'image contenant un visage .....	66
Figure 2.17 : Extraction du visage.....	66
Figure 2.18 : Interface du système développé .....	67
Figure 2.19 : Cas d'un document ou figure un visage d'enfant non détecté .....	69
Figure 3.1 : Système élaboré en utilisant la première approche proposée .....	72

Figure 3.2 : Phase d'apprentissage du réseau de neurones.....	73
Figure 3.3 : Différents exemples de documents traités par la première approche proposée .....	75
Figure 3.4 : Base de données utilisée pour les tests .....	76
Figure 3.5 : Système élaboré en utilisant la deuxième approche proposée .....	78
Figure 3.6 : Extraction des différentes positions exploitées par notre système.....	79
Figure 3.7 : Extrait de la base des données utilisée.....	82

## Liste des Tableaux

Tableau 2.1 : Etapes d'Extraction de l'image du document.....	58
Tableau 3.1 : Résultats Expérimentaux.....	76
Tableau 3.2 : Caractéristiques du réseau de neurones utilisé par notre système .....	81
Tableau 3.3 : Résultats expérimentaux concernant la deuxième version du système .....	83

# **Introduction**

La biométrie se présente comme une technologie potentiellement puissante afin de répondre à ces besoins liés à la sécurité, elle signifie l'étude quantitative des êtres vivants, autrement dit c'est une technique d'identification d'une personne au moyen de sa morphologie : son œil, son empreinte digitale, sa voix, son visage etc... , ou bien au moyen de son comportement : sa signature, sa façon d'écrire etc...

Depuis le début du XXI<sup>e</sup> siècle, le mot « biométrie » est utilisé dans un sens plus restrictif de l'identification des personnes en fonction de caractéristiques biologiques telles que les empreintes digitales, les traits du visage, etc.

La biométrie regroupe l'ensemble des techniques informatiques visant à reconnaître automatiquement un individu à partir de ses caractéristiques physiques, biologiques, voire comportementales. Les données biométriques sont des données à caractère personnel car elles permettent d'identifier une personne. La plupart de ces données ont la particularité d'être uniques et permanentes (ADN, empreintes digitales, etc.). Elles se rapprochent ainsi de ce qui pourrait être défini comme un « identificateur unique universel », permettant le traçage généralisé des individus.

C'est au début des années 1970 que les premières études théoriques sur la reconnaissance faciale sont apparues . Le passage vers des produits commerciaux n'a pas reçu une impulsion décisive qu'à partir des années 1994-1996 en bonne partie grâce à la mise en œuvre d'un programme d'évaluation international FERET (Face Recognition Technology), organisé par le ministère de la Défense américaine (Department of Defense, DoD). Une base d'images de référence est mise en disponibilité pour les tests. La fin du programme en 1996 a marqué l'apparition sur le marché des premiers produits commerciaux. Le DoD décida alors de lancer des programmes d'évaluation, FRVT (Facial Recognition Vendor Test) successivement en

2000, 2002 et en 2006 dont l'objectif était d'évaluer les performances des produits commerciaux [Cha 09].

## **Motivation et Problématique**

Au cours de ces dernières années, la sécurité est devenue une préoccupation internationale. La nécessité de la protection civile d'une part et la lutte contre les fraudes et les crimes d'autre part, placent au centre un dispositif sécuritaire pour de nombreux domaines comme par exemple le transport, le secteur bancaire, les services publics, et en particulier les services qui demandent une pièce d'identité (passeport, carte de séjour, etc.), ou un document où figure une photo du visage de la personne à identifier pour autoriser son accès à un tel service. Le dénominateur commun, est d'offrir des moyens simples, pratiques, fiables, pour vérifier ou identifier une personne, sans l'assistance d'une autre personne.

Dans cette thèse, nous nous intéressons à l'exploitation des informations en provenance d'un document mixte supposé contenant du texte et des images où figurent des visages. Le but de l'étude est d'être capable à reconnaître les visages situés dans ce document. Pour ce faire, plusieurs critères doivent être respectés :

- L'objet recherché – une personne – dont le visage est sujet à de nombreuses variations ; il faut prendre en compte non seulement les changements extrinsèques provoqués par la personne (conditions d'illumination, expression faciale, position) mais aussi les changements intrinsèques de la classe visage (âge, moustaches, barbe, cheveux, etc.) ;

- La détection et l'extraction du visage à partir du document donné en entrée envoie une image représentant une large scène dans le document ; bien que cette image soit de haute résolution, les objets recherchés à l'intérieur - les visages – ne sont représentés que par quelques centaines de pixels ; il faut évaluer les performances des méthodes pour de faibles résolutions;

– Le but visé est une application en temps réel ce qui impose des choix au niveau du temps d'exécution des méthodes utilisées : il faut que la méthode développée soit capable de détecter les visages dans le plus vite possible avant qu'un tel incident arrive; Les différentes tâches réalisées - segmentation, description, classification - devront donc être peu coûteuses en temps de calcul au moment de l'exécution.

Il existe, dans la littérature, des systèmes qui répondent d'une certaine mesure à ce cahier de charge. L'objectif de ce travail est donc de tester et de comparer les méthodes existantes, de proposer une combinaison judicieuse de ces méthodes, et d'apporter de nouvelles solutions pour améliorer l'existant. Le but à atteindre est un système complet avec :

- ☞ Une segmentation d'images puis celle du visage rapide et judicieuse permettant de réduire le champ de recherche du visage éventuel ;
- ☞ Un descripteur d'image rapide mais suffisamment pertinent pour permettre une classification correcte ;
- ☞ Un classifieur donnant une réponse rapide quant à la classe de l'objet, tout en ayant de bonnes performances de reconnaissance de visage et peu de fausses détections.

La rapidité se mesure en fait pour tout le système : il faut que l'exécution de la totalité des processus se fasse en «temps réel». Ici, cette expression sous-entend que le système soit capable de détecter et de reconnaître les visages présents dans le document avant qu'il arrive quoi que ce soit, afin de prendre une telle décision qui peut aboutir à un danger sur la population.

La performance de la globalité du système développé dépend des résultats fournis en sortie de la chaîne de traitement. Cela correspond donc aux résultats du classifieur et de la bonne reconnaissance du visage. Le système sera parfait s'il est capable de reconnaître tous les visages tout en ne donnant aucune fausse détection. En pratique, le

système recherché doit avoir un bon taux de reconnaissance avec un faible taux de fausses détections.

### **Architecture du système proposé**

Notre système s'articule sur les étapes suivantes, comme c'est illustré dans (Figure 1) présentée par la suite:

- L'étape de détection (à partir d'un document texte) de l'image contenant le visage à identifier : correspond à éliminer, d'une part, tous ce qui est non image (le texte du document) et d'autre part à éliminer tout ce qui est image ne contenant pas de visage. C'est une étape de classification où on assigne l'image à la classe visage ou à la classe non visage.
- L'étape de détection du visage contenu dans l'image identifiée : identifie et localise le visage dans l'image détectée dans l'étape 1, indépendamment de la position, de l'échelle, de l'orientation et de l'éclairage.
- L'étape d'extraction des caractéristiques du visage détecté : sert à extraire des informations pertinentes et discriminantes d'un visage. Les caractéristiques extraites sont des informations utiles à la phase de reconnaissance et elles sont modélisées pour fournir la signature biométrique du visage qui est ensuite utilisée dans la phase de classification.
- L'étape d'apprentissage consiste à traiter les représentations dans la phase d'extraction de caractéristiques pour le descripteur. Généralement, les deux étapes d'extraction des caractéristiques et d'apprentissage sont regroupées en une seule étape.
- L'étape de la classification de visages, dont on distingue deux tâches : l'identification et la vérification. En mode identification de visage, l'image à l'entrée du système est l'image d'un inconnu et le système doit rechercher l'identité de cet inconnu par comparaison de ses caractéristiques faciales à



l'ensemble des caractéristiques faciales des visages de la base de données d'individus connus.

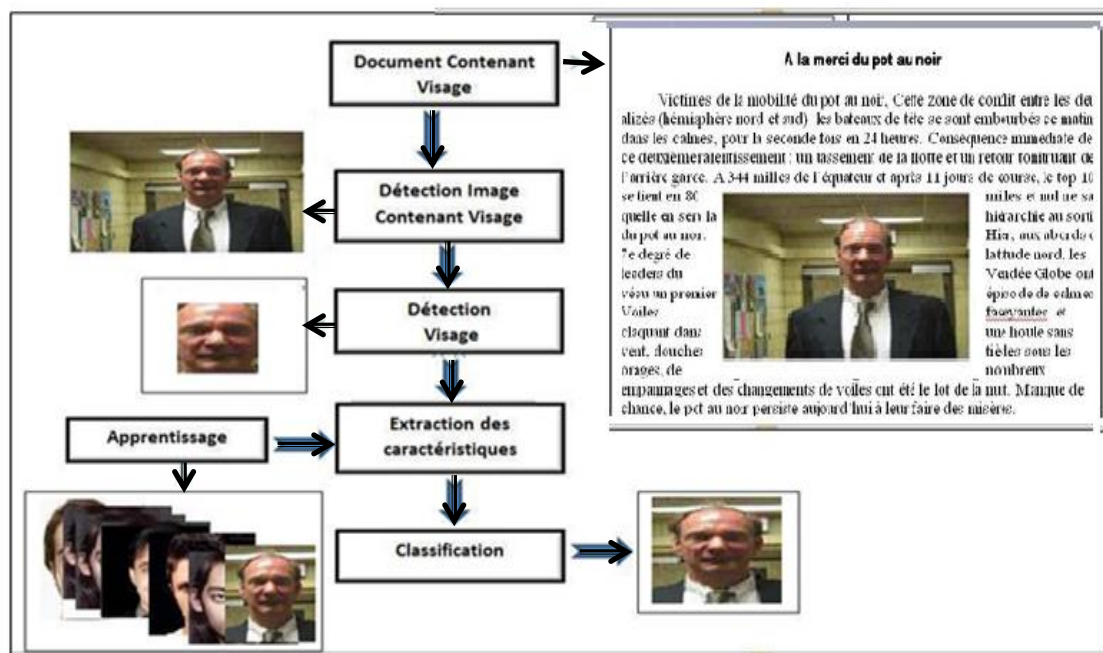


Figure 1- Architecture de système de reconnaissance de visage dans un document mixte.

Différents facteurs peuvent être considérés comme problèmes majeurs qui peuvent confronter la reconnaissance faciale à savoir les problèmes liés au changement de pose du visage, aux variations, d'illumination, d'expression et d'âge et au occultation partielle du visage. Ces variations sont les plus difficiles parce que les variations de l'apparence du visage d'une personne en conditions différentes de pose ou d'illumination sont souvent beaucoup plus importantes que les variations entre les images, de visage de deux individus différents, acquises dans les mêmes conditions.

## Contributions et plan du manuscrit

Dans le cadre de cette thèse, les travaux effectués portent sur l'identification automatique de visages dans un document mixte qui est composé à la fois du texte et des images ou figurent des visages. L'objectif est de simplifier le déroulement de l'identification biométrique qui consiste à la recherche de l'identité d'un individu parmi l'ensemble des individus de la base de données préenregistrée, appelée aussi

galerie. En effet, le schéma d'identification classique est complexe et très coûteux, notamment dans le cas de grandes bases de données biométriques. Ainsi nous proposons de réduire le taux d'identification, comme objectif essentielle, et d'améliorer les performances en termes de temps de calcul, comme second objectif.

Le manuscrit de cette thèse de doctorat s'articule autour de trois chapitres qui sont organisés de la manière suivante :

- Dans le premier chapitre nous donnons, dans sa première partie, une présentation générale de la biométrie. Elle présente une variété de définitions sur la biométrie et décrit l'architecture d'un système biométrique, puis définit, d'une manière générale, les différents modules intervenant dans la chaîne de traitement du système biométrique. Ensuite, la place de la reconnaissance faciale parmi les autres techniques biométriques est analysée. La seconde partie de ce chapitre, est consacrée à une analyse détaillée des différentes techniques développées au cours de ces dernières années dans les domaines de la détection et de la reconnaissance de visage. Notre objectif est de classer ces méthodes d'une manière efficace afin de mettre en évidence les particularités ainsi que les avantages et les inconvénients de chacune d'entre elles. Ceci nous permettra, par la suite, de mettre en avant les critères de performance qui vont guider le choix des solutions retenues dans le cadre de notre problématique.
- Dans le deuxième chapitre nous proposons un algorithme pour la détection d'un visage dans un document mixte. Cet algorithme s'articule sur les étapes suivantes :
  - Acquisition du document mixte ;
  - A l'aide d'une méthode basée sur la couleur de peau (Méthode Hybride), détecter, parmi toutes les figures y présent, celles ayant la même couleur que la couleur de la peau ;
  - Et parmi les figures détectées dans l'étape précédente, utilisation de la méthode Template Matching pour détecter, uniquement, les figures présentant un visage

humain, les quelles seront, ensuite, traitées pour réaliser une segmentation sur ces figures afin d'identifier les visages y.

- Le troisième chapitre est réservé à la présentation de l'algorithme proposé pour la reconnaissance du visage détecté dans la phase de détection, en exploitant une méthode géométrique, après avoir déterminé les caractéristiques du visage détecté. Nous évoquerons aussi à une comparaison, à partir des résultats expérimentaux déduits durant notre recherche, des différentes méthodes aussi bien en phase de détection qu'en phase de reconnaissance.

En conclusion, nous récapitulerons les principales contributions de ce travail avant d'exposer les perspectives envisagées.

# **1. Généralités sur la reconnaissance de visage**

## **1.1. Introduction**

Dans la communauté du traitement d'images et de la vision par ordinateur, l'analyse du visage est une discipline en plein essor. Nous définissons ici, dans ce chapitre, dans un premier temps les différentes catégories des systèmes d'authentification existantes faisant partie du domaine général de l'analyse du visage et nous donnons, également, l'architecture d'un système biométrique en indiquant, dans la mesure du possible, les références bibliographiques importantes à partir desquelles nous avons inspirés cette architecture.

Dans un deuxième temps, nous nous focalisons, d'abord, sur une description générale sur les différents formalismes qui ont été utilisés dans la littérature concernant la phase de détection d'un visage, puis nous nous intéresserons d'avantages sur la description détaillée des approches liées à cette phase utilisées par notre système développé.

Nous présentons, ensuite, les différentes méthodes d'analyse de reconnaissance de visages et plus spécifiquement celles utilisées, également, par notre système dont le but d'avoir une idée explicite sur la technique que nous avons abordé pour la reconnaissance de visages.

## **1.2. Le Contrôle d'accès par biométrie**

Parmi les problèmes d'actualité, le fait de savoir déterminer l'identité d'une personne de manière automatique. Les mots de passe et cartes d'identité ne peuvent pas fournir des fonctions d'authentications vitales comme la détection d'inscriptions multiples. Par exemple, les utilisateurs peuvent facilement nier l'utilisation d'un service en prétendant que leur mot de passe a été volé ou deviné. Les particuliers peuvent aussi cacher leur véritable identité en présentant des duplicatas de documents d'identité falsifiés.

Par conséquent, il devient de plus en plus évident que ces mécanismes ne sont pas suffisants pour déterminer d'une manière fiable l'identité d'une personne et qu'un mécanisme plus solide pour l'identification basé sur quelque chose que vous êtes, à savoir la biométrie, est plus que nécessaire.

Le mot biométrie a été défini comme l'étude mathématique des variations biologiques à l'intérieur d'un groupe déterminé.

La biométrie est une alternative aux anciens modes d'identification. Elle consiste à identifier une personne à partir de ses caractéristiques physiques ou comportementales [Jai 04]. Le visage, les empreintes digitales, l'iris, etc. sont des exemples de caractéristiques physiques. La voix, l'écriture, le rythme de frappe sur un clavier, etc., sont des caractéristiques comportementales. Ces caractéristiques, qu'elles soient innées comme les empreintes digitales ou bien acquises comme la signature, sont attachées à chaque individu et ne souffrent donc pas des faiblesses des méthodes basées sur une connaissance ou une possession [Per 02].

Donc avant de reconnaître n'importe quelle personne, il est indispensable d'identifier divers caractéristiques variantes d'une personne à une autre. Ce pendant le fait de localiser son visage, par exemple, est un vecteur essentiel de son identification. Cadre de notre thèse, la reconnaissance de visages est l'une des modalités biométriques couramment utilisée. Nous en ferons une présentation plus détaillée dans le but de préciser les enjeux et les motivations de cette thèse.

### **1.2.1. Taches de l'analyse du visage**

Au sein de notre société, l'identification d'une personne occupe, aujourd'hui, une place prépondérante. En effet le fait d'assurer l'identité d'une personne est nécessaire dans de nombreux domaines : contrôle d'accès à des zones réglementées, passage aux frontières, opérations bancaires, . . .

Et pour garantir l'identité d'un individu en limitant les risques d'usurpation, de nombreux systèmes d'authentification sont mis en place dans la vie quotidienne. Traditionnellement, ces systèmes sont classés en trois catégories [Chu 15] :

La première catégorie : l'identification est réalisée, par exemple, grâce à un mot de passe, un code PIN (Personal Identification Number). Ces informations, enregistrées par le fournisseur de service lors du premier contact avec l'utilisateur, permettent ensuite de s'assurer que l'individu est bien celui que l'on pense être. Ce premier type de méthode semble bien être accepté malgré la présence d'un certains nombres de limitation à titre d'exemple, un oubli du mot de passe qui peut perturber l'utilisateur lors de la demande de son authentification, ou vol du mot passe qui peut rendre l'accès au service non sécurisé.

En deuxième lieu, La catégorie des systèmes basés sur l'utilisation d'un objet pour l'accès à un service à titre d'exemple l'utilisation d'une clef ou un badge. Ce type de méthode semble, également, bien être accepté malgré la présence d'un certains nombres de limitation, à titre d'exemple limite principale, le risque de duplication de l'objet qui peut aboutir à une authentification multi-utilisateur au service.

Et comme troisième catégorie, le type d'authentification qui est basé sur les caractéristiques biométriques d'un individu (empreintes digitales, visage, iris, . . .). Ces caractéristiques lui permettent à l'utilisateur d'être authentifié, mais elles ont la particularité d'être, d'une part, universelles, du fait que tout le monde les possède, d'autre part, uniques, du fait que deux personnes ne peuvent pas partager une même caractéristique biométrique, et en fin, permanentes, du fait qu'elles ne varient pas à court terme. Ces méthodes permettent de garantir un niveau de sécurité plus élevé du fait que les données biométriques sont beaucoup plus difficiles à falsifier en comparaison avec les deux premières catégories.

La reconnaissance faciale est une technique commune et populaire. Elle est la plus acceptable puisqu'elle correspond à ce que les humains utilisent dans l'interaction visuelle. En la comparant aux autres méthodes, la reconnaissance faciale s'avère

plus avantageuse du fait que c'est une méthode non intrusive, c'est-à-dire elle n'exige pas la coopération du sujet (en observant les individus à distance).

### **1.2.2. Les composantes d'un système biométrique**

Un système biométrique est composé de trois modules [Lam 16] [Liu 01] [Jai 00]:

**1<sup>er</sup> module :** l'information biométrique, selon la modalité considérée pour identifier l'individu, est enregistrée grâce à un capteur. Pendant cette phase, la caractéristique biométrique est mesurée grâce à un capteur, on parle d'acquisition ou de capture. En général, cette capture n'est pas directement stockée et des transformations lui sont appliquées. En effet, le signal contient de l'information inutile à la reconnaissance et seuls les paramètres pertinents sont extraits. Le modèle est une représentation compacte du signal qui permet de faciliter la phase de reconnaissance, mais aussi de diminuer la quantité de données à stocker. Le processus d'adaptation est une étape qui s'avère nécessaire à ce module, il permet de maintenir et d'améliorer le système biométrique. En effet, au cours de la phase d'apprentissage, le système utilise quelques instances d'un même attribut ce qui ne reflète pas un modèle générale qui décrit toutes les variations possibles de cet individu. L'adaptation est, donc, indispensable pour rendre, pertinentes, les caractéristiques de l'attribut [Fre 00].

**2<sup>ème</sup> module :** des étapes de prétraitements et de post-traitements sont souvent appliquées avant et après l'extraction des descripteurs. Les prétraitements ont comme objectif d'améliorer la qualité de l'information biométrique initiale (amélioration de la qualité des images, par exemple), et les post-traitements ont souvent comme objectif de supprimer toute information non pertinente, telle que des faux descripteurs extraits de l'information brute (agissant comme du bruit). L'ensemble des descripteurs forme alors le gabarit biométrique, qui peut être stocké ou bien directement exploité lors de l'étape d'authentification (identification ou vérification).

**3<sup>ème</sup> module :** l'information biométrique est comparée à une autre information lors de l'étape du matching, et une décision est prise sur l'identité de l'individu en fonction du score retourné. Pendant cette phase, la caractéristique biométrique est mesurée et un ensemble de paramètres est extrait. Le capteur utilisé doit avoir des propriétés aussi proches que possibles du capteur utilisé durant la première. Si les deux capteurs ont des propriétés trop différentes, il faudra en général appliquer une série de prétraitements supplémentaires pour limiter la dégradation des performances. La suite de la reconnaissance sera différente suivant le mode opératoire du système.

### **1.3. Evaluation des performances d'un système**

Parmi les techniques biométriques, on distingue la reconnaissance faciale. Dans la vie quotidienne chacun de nous identifie, tout au long de la journée, différents visages. Ainsi lorsque nous rencontrons une personne, notre cerveau va chercher dans notre mémoire et vérifier si cette personne est répertoriée ou non.

Pour la reconnaissance de visage 2D, plusieurs méthodes ont été développées. Cependant, elles présentent un certain nombre de limitations liées à l'orientation du visage ou à la pose, à l'éclairage, à l'expression faciale, etc. Ces dernières années, on parle de plus en plus des techniques de reconnaissance de visage 3D comme une solution alternative pour résoudre les problèmes cités ci-dessus. En effet, la richesse de l'information fournie par les mesures 3D permet de reconstruire la forme tridimensionnelle du visage. Ce type de représentation du visage est invariant aux changements d'illumination et de pose [Gue 08].

Un système est jugé plus performant si le critère taux de reconnaissance est amélioré et le critère taux d'erreur du classifieur est réduit. Le premier critère indique le pourcentage d'individus correctement reconnus sur la totalité des images testées, le second critère est le pourcentage d'erreurs de reconnaissance commises sur toute la base. L'interprétation de ces approches d'évaluation varie selon l'application du système de reconnaissance de visages qui peut être utilisé selon deux objectifs :



l'identification d'individu qui signifie que le système doit fournir l'identité d'un individu à partir de son image faciale, ou la vérification d'identité qui signifie que le système vérifie la compatibilité d'une identité donnée par l'utilisateur et son image faciale captée à l'instant. La différence entre ces deux objectifs réside dans le fait que le premier objectif demande un parcours complet de la base de données de signatures, tandis que le deuxième objectif n'a pas besoin dans ce cas de parcourir toute la base de signatures, mais plutôt utilise des critères de validation définis a priori. Toutefois des seuils d'acceptation et de rejet sont généralement utilisés dans le premier cas pour délivrer la première décision : connue ou inconnue. Si c'est connu, le système s'appuie sur une technique de classification pour déterminer l'identité.

## **1.4. Bases des données disponibles**

Parmi les bases de données les plus reconnues dans la littérature concernant la reconnaissance des visages on cite les bases : FERET, AT&T, YaleFaces et AR.

Dans le cadre du programme de technologie de reconnaissance faciale (FERET), une base de données d'imagerie faciale a été collectée entre décembre 1993 et août 1996. La base de données est utilisée pour développer, tester et évaluer des algorithmes de reconnaissance faciale. L'acquisition des images a été opérée au sein de l'Université George Masson(USA) dans un environnement semi-contrôlé et durant plusieurs années. Cette base est considérée parmi les plus anciennes et les plus volumineuses, du fait qu'elle regroupe 14.126 images de 1.199 individus [Cha 09].

Dans le cadre d'un projet de reconnaissance faciale réalisé en collaboration avec le groupe Speech, Vision and Robotics du département d'ingénierie de l'Université de Cambridge, la base de données de visages AT&T, (anciennement «La base de données ORL des visages»), contient une série d'images de visage prises entre avril 1992 et avril 1994. Cette base de données contient dix images différentes de chacun des 40 sujets distincts. Pour certains sujets, les images ont été prises à des moments différents,

variant l'éclairage, les expressions faciales (yeux ouverts / fermés, souriant / non souriant) et les détails du visage (avec lunettes / sans lunettes) [Haj 15].

La base de données YaleFaces est composée d'images de 15 personnes, pour chacun sont stockées 11 images en niveaux de gris, soit 165 images au total. L'acquisition des images a été opérée au centre CVC (Center for Computational Vision and Control) de l'université Yale dans un environnement entièrement contrôlé. Les images contiennent des variations de luminosité et d'expressions faciales, elles ont été acquises toutes sur un fond blanc et avec une résolution de 320×243 [Cha 09].

C'est dans le centre CVC (Computer Vision Center) à l'Université Autonome de Barcelone que la base de données de visages AR[MB98] [Cha 09] a été créé. Elle contient 3.120 images en couleur correspondant aux visages de 120 personnes (65 hommes et 55 femmes). Elles ont été acquises sur un fond et une résolution de (255×255) qui étaient fixes quelles que soient les conditions de prise de vue.

## **1.5. Détection de visages : Etat de l'art**

Un système automatique de reconnaissance de visages se décompose en trois sous-systèmes : détection de visage, extraction des caractéristiques et reconnaissance de visages.

L'apparence globale du visage est en fonction des changements d'illumination, d'expression ou de pose. Les méthodes holistiques ne permettent donc pas une reconnaissance robuste à ces variations [Bap 15]. Les représentations locales des visages proposent de les caractériser par un ensemble de vecteurs caractéristiques extraits dans des zones précises du visage. Des approches consistent à capturer, pour chaque individu, une collection d'images acquises sous des poses variées. Les premiers travaux traitant de cette problématique ont été effectués par Beymer et al. [Bey 94].

La détection de visage est une tâche préliminaire nécessaire à la plupart des techniques d'analyse du visage. Elle consiste à déterminer la présence ou l'absence de visages

dans une image et en cas de présence à déterminer sa localisation. Les techniques utilisées sont généralement issues du domaine de la reconnaissance des formes.

La détection des visages pose le problème de la localisation des visages présents dans une image d'entrée. L'intérêt de la localisation faciale va au-delà de l'application de ce présent travail. Son utilité se manifeste dans des domaines variés allant de la vidéo surveillance au jeu interactif. Les premières difficultés rencontrées par les méthodes s'attendant à détecter les visages sont les variations de pose (vue de profile, de face), d'expression, de rotation du visage, d'âge et d'illumination. Ce dernier type de difficulté pouvant être surmonté par un prétraitement de normalisation et de compensation de l'illumination.

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, sont des paramètres qui peuvent être maîtrisés.

### **1.5.1. Classification des méthodes de localisation faciale**

Nous présentons dans cette section les différentes modélisations informatiques du visage, utilisées dans la littérature. Il peut s'agir de modèles qui servent aussi bien à l'analyse de l'identité qu'à l'analyse des expressions. Nous avons distingué deux types de méthodes : celles dites « globales », considérant le visage dans son ensemble sans traitement particulier pour certaines composantes et celles basées sur une segmentation explicite du visage en composantes et une description des caractéristiques de ces composantes faciales.

#### **1.5.1.1. Approche globale sans segmentation**

Cette approche consiste à prendre en considération un ensemble de points du visage représentant l'état de certaines composantes. Ces points doivent correspondre à des indices visuels qu'il est possible de mettre en correspondance sur toutes les observations de l'étude. Les points d'analyse diffèrent quand il s'agit d'analyser l'identité de quand il s'agit d'analyser l'expression.

Les coordonnées de chacun des points du modèle ainsi que la valeur des pixels en leur voisinage sont les éléments caractéristiques d'un visage. Il lui permette de définir un descripteur de visage plus puissant que l'image brute. Certaines méthodes considèrent un traitement particulier en chacun des points d'intérêt : le résultat du traitement en chacun des points formant le vecteur d'entrée du système d'analyse.

Comme c'est illustré dans la figure 1.1, nous donnons à titre d'exemple, [Mer 07] la transformée par ondelettes de Gabor, utilisée sur une grille de points aussi bien pour la reconnaissance d'identité. L'ensemble des réponses des filtres de Gabor forme un vecteur transmis en entrée à un système de reconnaissance. Ces vecteurs d'entrée sont présentés à un réseau de neurones pour la reconnaissance d'identité. Divers expériences ont témoigné que les points les plus importants pour la tâche de reconnaissance des émotions sont les points au tour des yeux, de la bouche, des sourcils et du menton.

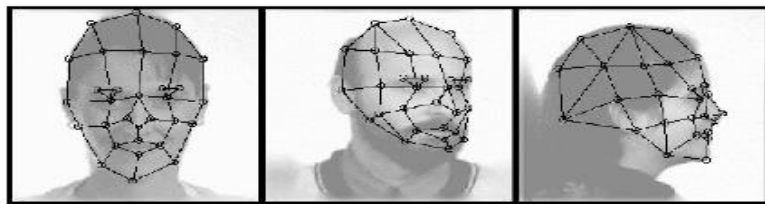


Figure 1.1 Modèles de visage basé sur un maillage de points pour la reconnaissance d'identité [Mer 07]

### **1.5.1.2. Approche avec segmentation**

Dans [Mer 07], Tian et al. [Li Tian 01] ont proposé un système d'analyse automatique des expressions faciales, par reconnaissance. Dans ce système les réseaux de neurones sont également utilisés dans la phase de classification. Un ensemble de mesures effectuées sur des descripteurs principalement géométriques du visage est considéré comme des données d'entrée des réseaux de neurones. Les sourcils et le haut des joues sont modélisés comme formes géométriques, d'un segment, et les yeux et la bouche sont modélisés comme courbes ayant des formes plus libres tenant compte de différents modèles de celles-ci ayant plusieurs états possibles : ouvert, semi-ouvert et fermé. En plus de ces descripteurs, la présence de rides est détectée par une analyse de contour dans certaines zones (haut du nez par exemple).

L'ensemble de ces données est fourni en entrée à des réseaux de neurones multicouches ayant une sortie avec un taux d'environ 90% et un taux de fausses alarmes d'environ 10%.

### **1.5.2. Méthode de détection de visage de Viola & Jones**

Les chercheurs, Paul Viola et Michael Jones, ont proposé la méthode Viola et Jones [Vio 01] qui est une méthode de détection d'objet dans une image numérique. Elle fait partie des toutes premières méthodes capables de détecter efficacement et en temps réel des objets dans une image. Inventée à l'origine pour détecter des visages, elle peut également être utilisée pour détecter d'autres types d'objets comme des voitures ou des avions. La méthode de Viola et Jones est l'une des méthodes les plus connues et les plus utilisées, en particulier pour la détection de visages et la détection de personnes.

La méthode de Viola & Jones est une méthode plus performante à l'heure actuelle. Elle permet de balayer une image, en utilisant une fenêtre de détection de taille initiale 24 px par 24 px et de déterminer si un visage y est présent. Lorsque l'image a été parcourue entièrement, la taille de la fenêtre est augmentée et le balayage recommence, jusqu'à ce que la fenêtre fasse la taille de l'image. L'augmentation de la taille de la fenêtre se fait par un facteur multiplicatif de 1,25.

Quant au balayage, il consiste à décaler la fenêtre d'un pixel. Ce décalage peut être changé afin d'accélérer le processus, mais un décalage d'un pixel assure une précision maximale. Ce qui la différencie des autres est [Vio 01]:

- l'utilisation d'images intégrales qui permettent de calculer plus rapidement les caractéristiques.
- la sélection par boosting des caractéristiques.
- la combinaison en cascade de classifieurs boostés, apportant un net gain de temps d'exécution.

La sélection par boosting consiste à remplacer un seul classifieur "fort" par plusieurs classifieurs "faibles" mis en cascade. Cependant, avec un seul classifieur dit "fort", qui se présenterait de la manière présentée dans la figure 1.2, Il faudrait attendre que le

classifieur analyse toute la fenêtre afin de savoir si un visage est présent dans l'image ou non.

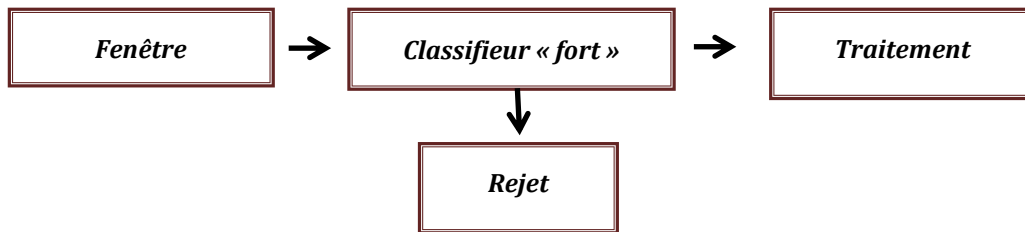


Figure I.2 : Sélection par un seul classifieur « fort »

En tant que procédé d'apprentissage supervisé, la méthode de Viola et Jones nécessite quelques centaines à plusieurs milliers d'exemples de l'objet que l'on souhaite détecter, pour entraîner un classifieur. Une fois son apprentissage réalisé, ce classifieur est utilisé pour détecter la présence éventuelle de l'objet dans une image en parcourant celle-ci de manière exhaustive, à toutes les positions et dans toutes les tailles possibles. Considérée comme étant l'une des plus importantes méthodes de détection d'objet, la méthode de Viola et Jones est notamment connue pour avoir introduit plusieurs notions reprises ensuite par de nombreux chercheurs en vision par ordinateur, à l'exemple de la notion d'image intégrale ou de la méthode de classification construite comme une cascade de classifieurs boostés.

### **1.5.2.1. Éléments de la méthode**

La célèbre technique de détection de visage dans la dernière décennie est celle proposée par Paul viola et Michael Jones. C'est une méthode pour la détection de visages dans une image numérique ou séquence vidéo puis utilisée pour détecter d'autres objets comme les voitures, etc.

Cette méthode combine quatre contributions clés :

- les caractéristiques pseudo-Haar ;
- l'approche d'image intégrale ;
- la méthode d'apprentissage adaptative AdaBoost ;

- l'algorithme en cascades de classifieurs.

Dans la suite, nous allons détailler ces étapes clés tout en précisant l'apport engendré de chacun d'entre elles sur la performance et l'efficacité de la méthode.

### **1.5.2.2. Caractéristiques pseudo-Haar**

Une caractéristique pseudo-Haar est représentée par un rectangle défini par son sommet, sa hauteur, sa longueur et son poids. Le recours au traitement d'images par ces caractéristiques est motivé par sa rapidité contre le traitement direct à travers le balayage de la totalité d'image pixel par pixel.

### **1.5.2.3. Approche d'image intégrale**

La méthode d'image intégrale est utilisée pour déterminer la présence ou l'absence des caractéristiques dans chaque position de l'image et à n'importe quelle taille. Le but de cette méthode à part la détection de caractéristiques est la réduction du temps de calcul de ces dernières. La valeur intégrale de chaque pixel est la somme de tous les pixels au-dessus de lui et de sa gauche.

### **1.5.2.4. Algorithme AdaBoost**

Paul Viola et Michael Jones ont utilisé l'algorithme AdaBoost pour sélectionner les caractéristiques de Haar à utiliser et pour fixer le niveau du seuil adéquat pour cette sélection. AdaBoost combine plusieurs classifieurs « faible » issus des caractéristiques pseudo-Haar pour former un classifieur « fort ». AdaBoost sélectionne un ensemble de classifieurs faibles où chacun « pousse » la réponse finale un degré vers la bonne direction. Il combine ces derniers et assigne un poids à chacun.

### **1.5.2.5. Algorithme en cascades de classifieurs**

La quatrième contribution de la méthode « Viola et Jones » est la détection en cascades. Une cascade se compose de  $n$  filtres dont chacun est un classifieur « faible » composé d'une seule caractéristique pseudo-Haar. Au cours d'une détection, si un filtre échoue à passer une sous-région alors elle est immédiatement classée comme « Non visage » sinon la région est passée vers le filtre suivant. Les sous-régions de l'image

qui traversent la totalité de cascade sont classées comme « visage » et tous les autres sont classées « Non Visage ».

Les poids qu'AdaBoost attribue aux filtres déterminent l'ordre des filtres dans la cascade commençant par le poids le plus lourd vers celui le plus faible pour éliminer les régions « Non visage » le plus tôt possible.

Une mise en cascade de classifieurs dont le critère de sélection serait moins sévère se présenterait dans la figure 1.3 présentée ci-dessous :

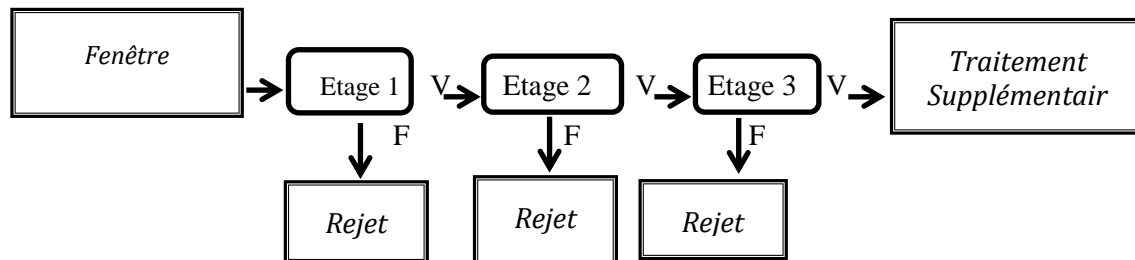


Figure 1.3 : Sélection par boosting

Par conséquent, dès que l'un des étages estime qu'il n'y a pas de visage, la fenêtre est rejetée et l'algorithme passe à la suite ce qui permet un gain de temps considérable.

### 1.5.2.6. Étapes clés de la méthode.

L'apprentissage est réalisé sur un très large ensemble d'images positives (c'est-à-dire contenant l'objet) et négatives (ne contenant pas l'objet). Plusieurs milliers d'exemples sont en général nécessaires. Cet apprentissage comprend :

- Le calcul des caractéristiques pseudo-Haar sur les exemples positifs et négatifs ;
- L'entraînement de la cascade : à chaque étage de la cascade, un classifieur fort est entraîné par AdaBoost. Il est construit par ajout successif de classifieurs faibles entraînés sur une seule caractéristique, jusqu'à l'obtention de performances conformes aux taux de détection et de fausse alarme souhaités pour l'étage.

La détection s'applique sur une image de test, dans laquelle on souhaite déceler la présence et la localisation d'un objet. En voici les étapes :



- Parcours de l'ensemble de l'image à toutes les positions et échelles, avec une fenêtre de taille  $24 \times 24$  pixels, et application de la cascade à chaque sous-fenêtre, en commençant par le premier étage ;
- Calcul des caractéristiques pseudo-Haar utilisées par le classifieur de l'étage courant, puis le calcul de la réponse du classifieur,
- Passage ensuite à l'étage supérieur si la réponse est positive, à la sous-fenêtre suivante sinon, et enfin l'exemple est déclaré positif si tous les étages répondent positivement ;
- Fusion des détections multiples : l'objet peut en effet générer plusieurs détections à différentes positions et échelles ; cette dernière étape fusionne les détections qui se chevauchent pour ne retourner qu'un seul résultat.

Dans le cadre de cette thèse, nous nous intéressons en particulier, d'une part, aux techniques de détection de visages basées sur l'analyse de la couleur de la peau « Hybride Methode » et d'autre part, à une technique par appariement du gabari « Template Matching ». Dans ce qui suit nous présenterons une étude détaillée sur chacune de ces techniques, ce qui nous permettra de situer notre approche par rapport aux travaux existants.

### **1.5.3. Méthode de détection de visage basée sur l'analyse de la couleur de la peau**

Ce type de méthode est considéré efficace et rapide [Ben 07]. Il réduit l'espace de recherche de la région visage dans l'image. De plus, plusieurs facteurs peuvent influencer sur la couleur de la peau, du fait qu'elle est une information robuste, tels que les rotations, les changements d'échelle, et les occultations partielles. L'efficacité de détecter, dans l'image, les pixels qui ont la couleur de la peau, dépend essentiellement de l'espace couleur choisi. Divers espaces couleur peuvent être utilisés pour détecter, dans l'image, les pixels qui ont la couleur de la peau, parmi les espaces les plus utilisés [Gue 08] :

- L'espace RVB, mis au point par la Commission Internationale de l'Eclairage (CIE) en 1931. Sa caractéristique consiste à représenter l'espace des couleurs à partir de trois rayonnements monochromatiques de couleurs : Rouge-Vert-Bleu.
- L'espace HSL (Hue, Saturation, Luminance), on l'appelle également l'espace TSL (Teinte, Saturation, Luminance) en Français, il est proche de la perception physiologique de la couleur par l'œil humain. C'est un modèle de représentation dit "naturel". Le modèle HSL consiste à décomposer la couleur selon des critères physiologiques :
  - la teinte (en Anglais Hue), correspondant à la perception de la couleur.
  - la saturation, décrivant la pureté de la couleur, c'est-à-dire son caractère vif ou terne.
  - la luminance, indiquant la quantité de lumière de la couleur, c'est-à-dire son aspect clair ou sombre.

Toutefois il existe d'autres modèles naturels de représentation proches du modèle HSL qui se distinguent soit par la brillance, soit par la valeur ou par l'intensité, parmi ces modèles :

- HSB : Hue, Saturation, Brightness soit en Français Teinte, Saturation, Brillance.
- HSV : Hue, Saturation, Value soit en Français Teinte, Saturation, Valeur.
- HSI : Hue, Saturation, Intensity soit en Français Teinte, Saturation, Intensité.
- YCrCb : codé à partir du signal RVB dont le paramètre Y représente l'information en noir et blanc et les paramètres Cr et Cb représentent l'information couleur.
- CIELab connue par  $L^*a^*b^*$  caractérisé par le fait que la couleur est repérée par trois valeurs :
  - ☞  $L^*$ , la luminance, exprimée en pourcentage (0 pour le noir à 100 pour le blanc).

---

☞  $a^*$  et  $b^*$  deux gammes de couleur allant, respectivement, du vert au rouge et du bleu au jaune avec des valeurs allant de -120 à +120.

Parmi les techniques de détection du visage basées sur la couleur de la peau, nous allons maintenant aborder deux types de méthodes : les méthodes explicites et les méthodes non paramétriques. Toute fois il existe d'autres méthodes telles que les méthodes paramétriques et les méthodes semi-paramétriques. Le point commun entre toutes ces approches est le fait qu'elles pratiquent une phase d'apprentissage sur un nombre d'images représentatives pour calculer une densité de probabilité de la couleur de la peau.

### **1.5.3.1. Les méthodes explicites**

Pour la détection des pixels ayant la couleur de la peau, les méthodes explicites utilisent des règles de décision empiriques et/ou statistiques [Gue 08]. Une méthode explicite est une méthode de classification qui consiste à définir explicitement les frontières de la région peau (cluster) dans l'espace couleur utilisé. Dans [Ben 07], [Gue 08], [Pee 03.], par exemple, la peau est classifiée dans l'espace couleur RGB en utilisant les règles suivantes :

$$\begin{aligned} R > 95 \text{ et } G > 40 \text{ et } B > 20 \text{ et} \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 \text{ et} \\ \text{abs}(R - G) > 15 \text{ et } R > G \text{ et } R > B \end{aligned} \quad (1)$$

Le point fort de ces méthodes est la simplicité des règles de détection de la peau qu'elles utilisent, ce qui engendre une classification rapide. Mais, leur problème principal est la difficulté de déterminer empiriquement un espace couleur approprié et d'impliquer des règles de décision adéquates qui assurent un taux de reconnaissance élevé.

Pour résoudre ce problème, une méthode utilisant des algorithmes d'apprentissage a été proposée [Gom 02]. Un espace RGB normalisé a été choisi  $r + g + b = 1$  (où  $r = R/(R+B+G)$ ,  $g = G/(R+B+G)$  et  $b = B/(R+B+G)$ ) étant la somme des trois composantes normalisées) sur lequel ils appliquent un algorithme d'induction

constructive afin de créer de nouveaux ensembles d'attributs pour les composantes RGB. Une règle de décision, semblable à l'équation (1) qui réalise la meilleure identification possible, par conséquent cette méthode a obtenu des résultats meilleurs que ceux qui sont obtenus avec un classifieur de Bayes défini dans l'espace RGB.

### **1.5.3.2. Les approches non paramétriques**

Pour représenter la distribution (ou densité de probabilité) de la tonalité de la peau dans un espace couleur, les approches non paramétriques sont caractérisées par l'utilisation des histogrammes couleur 2D et 3D. L'intérêt d'utiliser les histogrammes couleur est qu'ils sont robustes (invariants) aux occultations et aux changements de point de vue [Gue 08].

Le déroulement des approches non paramétriques se fait, en général, en trois étapes :

- Construction des histogrammes de couleur de peau et de non peau.
- Calcul de la probabilité conditionnelle pour chaque couleur de peau et de non peau.
- Calcul de la probabilité, pour qu'un pixel corresponde à la classe peau, se fait par la règle de Bayes.

Et pour extraire les régions peau, un seuil de classification est déterminé à partir de la relation entre les détections correctes et les détections fausses.

En conclusion, la plupart des méthodes de détection de la peau fournissent des résultats obtenus sur leurs propres bases de données qui ne sont pas disponibles. La base d'apprentissage et de test la plus utilisée pour la détection de la peau est la base de données Compaq [Jon 99].

### **1.5.4. Template Matching**

La technique de la méthode Template Matching, consiste à déterminer plusieurs sous régions de l'image en entrée puis de comparer l'intensité des pixels entre chaque région

à un gabarit prédéfini [Ben 07]. Ce processus consiste en pratique à effectuer plusieurs balayages couvrant toute la superficie de l'image dont chacun de ces balayage, une distance sera déterminée entre le gabarit et la sous-région de l'image à analyser ainsi la distance minimale de ces distances entraine l'endroit le plus propice à la présence de visage comme le montre la figure I.4.

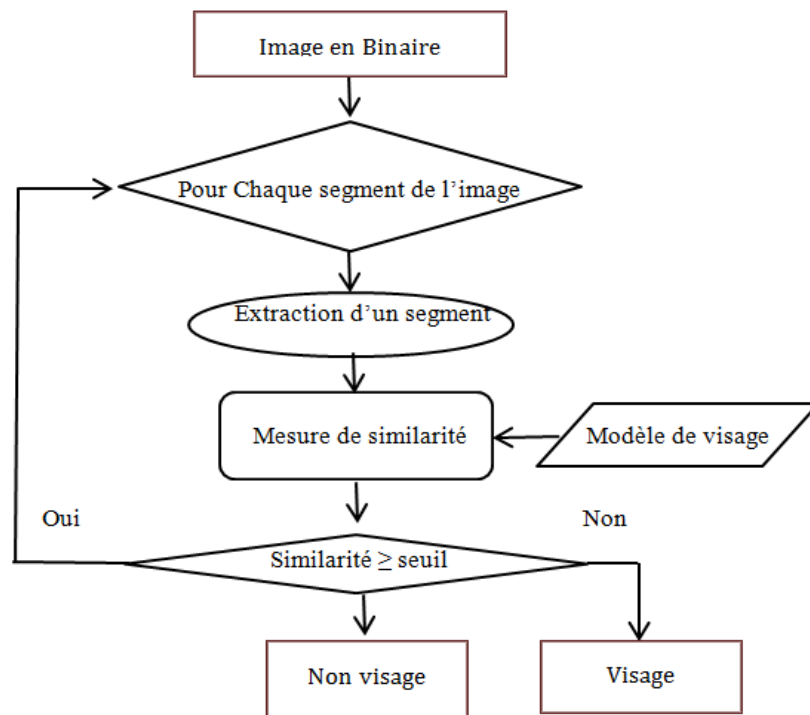


Figure I.4 : Schéma illustratif de la méthode Template Matching

Parmi les distances utilisées dans ce processus, il y a notamment la norme  $L_1$  (Distance de Manathan), la norme  $L_2$  (Distance Euclidienne), la cross-corrélation, ...etc.

Le choix d'un gabarit efficace représente un défi en lui-même. En effet, l'utilisation d'un gabarit plus ou moins adapté au type d'objet recherché peut nuire à une détection efficace et diminuer la précision des résultats.

Un modèle de visage peut être créé en effectuant un calcul de la moyenne à partir de plusieurs images de visages en niveaux de gris.

Le principe de cette méthode consiste à prendre une image donnée quelconque et un modèle d'image prédéfini supposé une sous-région dans l'image à analyser, puis de déterminer le degré d'appartenance (similarité) de ce modèle à cette région pour déduire quelle région est plus similaire au modèle. Comme c'est indiqué au paravent, diverses distances peuvent être utilisées pour mesurer la similarité entre deux images (le gabarit et la sous-région de l'image à analyser). Parmi lesquelles on cite la norme  $L_1$  qui correspond à une simple différence de norme donnée par la première formule pour la norme  $L_1$ , et la deuxième formule pour la norme  $L_2$  [Bou 16][Heu 02].

$$\text{Norme } L_1 : \quad \sum_{(i,j)} |a_{(i,j)} - b_{(i,j)}| \quad (2)$$

$$\text{Norme } L_2 : \quad \sqrt{\sum_{(i,j)} (a_{(i,j)}^2 - b_{(i,j)}^2)} \quad (3)$$

Où les  $a_{(i,j)}$  et  $b_{(i,j)}$  sont la valeur du pixel  $(i,j)$  dans les deux images à comparer. Il ressort donc que plus ces valeurs sont petites, plus les images sont proches.

## **1.6. Reconnaissance faciale : Etat de l'art**

L'apport de techniques automatisées performantes dans les systèmes de surveillance devient de plus en plus étendus dans les milieux urbanisés. Cette automatisation permet de répondre à l'accroissement du nombre de caméras présentes dans les systèmes tout en accommodant la quantité croissante d'individus impliqués dans l'environnement de capture.

Dans le but d'éviter les erreurs de classification des gens détectés, l'identification automatisée des individus peut s'accomplir par une reconnaissance de visages qui nécessite l'application de techniques avancées et robustes.

Afin d'améliorer les systèmes de sécurité avancés comme dans les aéroports, il devient possible d'employer des méthodes de reconnaissance de visages qui permettraient d'identifier rapidement des individus marqués dans une liste de contrôle.

Par contre, l'inscription de ces individus dans le système est souvent limitée par une quantité réduite d'images de référence. Dans la majorité des cas, seulement une image est disponible pour représenter un individu particulier, ce qui rend l'entraînement du système beaucoup plus complexe à cause de la limitation d'information.

La reconnaissance de visage consiste à associer une identité à un visage après l'avoir détecté. On distingue deux types différents de reconnaissance :

**1<sup>er</sup> type :** l'identification d'un individu, où il s'agit de trouver dans une base de données de visages d'individus différents, le visage le plus ressemblant à celui de l'individu étudié.

**2<sup>ème</sup> type :** l'authentification d'un individu, où il s'agit de vérifier que le visage de l'individu étudié a bien l'identité qu'il prétend posséder.

Ces deux manières de poser le problème font intervenir des opérateurs d'analyse très différents. Une analyse de performance de plusieurs algorithmes est alors nécessaire pour en retenir les meilleurs dont l'objectif est de profiter d'un système opérationnel et efficace. Ainsi, le présent projet vise à explorer plusieurs techniques formant l'état de l'art du domaine de la reconnaissance de visages afin d'en déduire des sources plus prometteuses pour le cas exploré. Plus particulièrement, des méthodes spécifiques de classifications ont été modélisées, puis sont progressivement améliorées, afin de tendre vers un système d'identification des individus ayant le plus grand taux d'exactitude. De plus, une quantité de temps notable est accordée à l'étude de performance des techniques d'extraction de caractéristiques de visages. Ainsi, il devient possible d'intégrer simultanément des descripteurs de visages plus représentatifs des individus de sorte à simplifier par la suite la tâche de classification et d'identification de ceux-ci.

### **1.6.1. Phase d'extraction du vecteur caractéristique du visage**

L'importance du module d'extraction de signatures est cruciale dans le processus de reconnaissance de visage. De ce fait, nous allons nous focaliser, dans ce qui suit sur la présentation des principales méthodes de modélisation paramétrique décrites dans la littérature scientifique. Nous allons présenter, au début, un formalisme général permettant de bien situer les méthodes les unes par rapport aux autres. Ce formalisme permet de distinguer les méthodes en fonction des atomes utilisés comme base de la modélisation.

Toutes les méthodes présentées possèdent des avantages et des inconvénients, ce qui explique qu'aucune n'occupe le premier rang devant toutes les autres.

Dans cette étape de modélisation, connue également sous le nom d'indexation, est extraite des informations d'image du visage qui peuvent modéliser le visage d'une personne par un vecteur de mesure qui le caractérise (vecteur caractéristique ou signature). Le choix judicieux de la méthode d'extraction est extrêmement important dans la mesure où la prochaine étape (classification) ne sera faite que sur la base des valeurs de ces vecteurs.

#### **1.6.1.1. La méthode d'Analyse en Composantes Principales(ACP).**

Le but de l'approche ACP est de capturer la variation dans une collection d'images de visages et d'utiliser cette information pour coder et comparer les visages (en termes mathématiques : trouver les vecteurs propres de la matrice de covariance de l'ensemble des images de visages). Le nombre possible de visages propres peut être approximé en utilisant seulement les meilleurs visages propres qui correspondent aux plus grandes valeurs propres [Tur 91]. Cette approche rencontre le problème du temps des calculs élevé et celui de la détermination du nombre de visages propres utiles [Dai 98].

L'algorithme ACP, connu sous le nom de "EigenFaces" (visages propres), adapté à l'analyse et l'identification de visage a été développé par M.A. Turk et A.P Pentland en



1991 [Tur 91] [Fic 12]. Il est composé de deux phases, une phase d'apprentissage et une phase de classification. Au cours de la première phase, un espace propre est construit à partir d'une base d'apprentissage en utilisant la méthode ACP puis ces mêmes images sont projetées sur l'espace ainsi obtenu. Durant la deuxième phase, un visage test est projeté à son tour sur ce même espace pour être alors identifié en le comparant aux projections de chacun des visages de la base d'apprentissage. On considère, alors, une image de taille  $N = n \times m$  comme un vecteur de dimension  $N$ . Autrement dit on représente ce visage comme un point dans un espace de dimension  $N$ . Soit  $\Gamma_i$  le vecteur de dimension  $N$  correspondant à une image  $I$  de la base d'apprentissage composée de  $M$  images. Soit  $\Psi$  l'image moyenne de cette base d'apprentissage définie par :

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (4)$$

Soit  $\Phi_i$  un vecteur de dimension  $N$  correspondant à l'image  $i$  dont on a soustrait l'image moyenne telle que :

$$\Phi_i = \Gamma_i - \Psi \quad (5)$$

La méthode ACP qui permet de trouver un ensemble de  $K$  vecteurs orthogonaux  $P_j$  décrivant de façon adaptée la distribution de ces visages en réalisant une combinaison linéaire des vecteurs de la base originale. Pour cela, on souhaite éliminer la redondance d'information traduite en termes mathématiques par la matrice de covariance  $C_x$  définie comme une matrice de dimension  $N \times N$  :

$$C_x = \sum_{i=1}^M \Phi_i \Phi_i^t = X X^t \quad (6)$$

Où  $X$  est la matrice de dimension  $N \times M$  telle que :  $X = [\Phi_1, \Phi_2, \dots, \Phi_M]$

Plus la valeur d'une variance est élevée plus elle traduit la différence existant entre deux images. Autrement dit, les directions associées aux variances élevées sont gardées et correspondent aux directions principales. Les directions principales ne sont rien d'autre que les  $M$  vecteurs propres  $V_i$  de la matrice  $A$  définie par :

$$A = XX^t. \quad (7)$$

Ensuite on projette les images sur l'espace des visages  $E_v$  ainsi formé.  $E_v$  est donné par la matrice:  $E_v = [X.v_1, X.v_2, \dots, X.v_{M'}]$ , la matrice de projection  $k$  qui décrit la  $K^{\text{ème}}$  classe de visage est définie par :  $\Omega_K^t = [\omega_{K_1}, \omega_{K_2}, \dots, \omega_{K_{M'}}]$ . Enfin, le projeté  $\Phi_p$  sur  $E_v$  d'une image  $\Phi$  (ramenée à sa moyenne) est donnée par :

$$\varphi_p = \sum_{i=1}^{M'} \omega_i X.v_i . \quad (8)$$

Pour l'identification, l'idée est de trouver le  $k^{\text{ème}}$  visage de la base d'apprentissage qui minimise la distance euclidienne  $\varepsilon$  entre le vecteur de projection  $\Omega_k$  et celui de l'image test  $\Omega_t$  :

$$\varepsilon = ||\Omega_t - \Omega_k||. \quad (9)$$

Cette distance peut être seuillée pour minimiser les erreurs de classification.

### **1.6.1.2. Moments de Hu**

Dans [Ela12], à partir des moments géométriques, [Hu 61] Hu a proposé un ensemble de sept moments invariants aux translations, rotations et changement d'échelle. Ils sont très utilisés dans la littérature pour la description de formes en vue d'une classification ou d'une indexation, mais sont assez sensibles aux bruits.

La formule du moment Hu  $m_{pq}$  d'une fonction de distribution  $f(x, y)$  est comme suit :

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad (10)$$

Pour une image digitale  $g(x, y)$  de taille  $M \times N$ , la formule ci-dessus devient :

$$m_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x^p y^q g(x, y) \quad (11)$$

Le moment central  $u_{pq}$  est donné par :

$$u_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} (x - \bar{x})^p (y - \bar{y})^q g(x, y) \text{ Avec } \bar{x} = \frac{m_{10}}{M_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \quad (12)$$

On définit le moment central normalisé :

$$n_{pq} = \frac{u_{pq}}{u_{00}^\gamma}, \quad \gamma = \frac{p+q}{2} + 1 \quad p + q = 2, 3, \dots \quad (13)$$

Hu a défini les sept moments, ci-dessous, qui caractérisent une image :

$$\emptyset_1 = n_{20} + n_{02} \quad (14)$$

$$\emptyset_2 = (n_{20} + n_{02})^2 + 4 n_{11}^2 \quad (15)$$

$$\emptyset_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2 \quad (16)$$

$$\emptyset_4 = (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2 \quad (17)$$

$$\emptyset_5 = (n_{30} - 3n_{12})(n_{30} + n_{12})[(n_{30} + n_{12})^2 + (n_{21} + n_{03})^2] + (3n_{21} - n_{03})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \quad (18)$$

$$\emptyset_6 = (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}) \quad (19)$$

$$\emptyset_7 = (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] + (n_{30} - 3n_{12})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \quad (20)$$

**1.6.1.3. Moments de Legendre**

Les moments de Legendre ont été présentés par Teague. [Bou 16] [Tea 80] Ils ont été utilisés dans de nombreuses applications de reconnaissance de formes [Cho 04]. Les moments de Legendre pour une image discrète de M x N pixels avec la fonction d'intensité f (x, y) ont été définis de la manière suivante [Ouj 12]:

$$L_{pq} = \lambda_{pq} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P_p(x_i)P_q(y_j)f(x, y) \tag{21}$$

Où :

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{M*N}$$

$x_i$  et  $y_j$  désignent les coordonnées de pixels normalisées dans l'intervalle [1, 1], qui sont données par:

$$\begin{cases} x_i = \frac{2x-(M-1)}{M-1} \\ y_j = \frac{2y-(N-1)}{N-1} \end{cases} \tag{22}$$

$P_p(x)$  est le  $p^{ième}$  ordre des polynômes de Legendre définie par :

$$P_p(x) = \sum_{k=0}^p \left\{ \frac{(-1)^{\frac{p-k}{2}} x^k (p+k)!}{2^p k! (\frac{p-k}{2})! (\frac{p+k}{2})!} \right\} \tag{23}$$

Et, la formule récurrente des polynômes de Legendre est:

$$\begin{cases} P_p(x) = \frac{(2p-1)x}{p} P_{p-1}(x) - \frac{(p-1)}{p} P_{p-2}(x) \\ P_1(x) = x, \quad P_0(x) = 1 \end{cases} \tag{24}$$

Dans ce travail, la formule récurrente est utilisée pour le calcul de polynômes de Legendre en vue d'augmenter la vitesse de calcul.

**1.6.1.4. Les Méthodes Géométriques**

Les méthodes géométriques consistent à extraire la position relative des différents éléments du visage (tels que le nez, la bouche et les yeux) afin de restreindre l'espace des données initiales. Différentes méthodes statistiques, inspirées des algorithmes classiques de reconnaissance de formes, sont alors utilisées pour discriminer les visages à partir de ces mesures [Ler 16]. Elles ont présenté un système d'identification de visage fondé sur l'utilisation des propriétés géométriques. Ce système se décompose, en premier lieu, sur une localisation d'éléments caractéristiques du visage, puis, en deuxième lieu, sur l'extraction de leur frontière et enfin sur une analyse factorielle des données ainsi extraites. La localisation des régions contenant les éléments significatifs du visage, tels que la bouche, les yeux et le nez, est effectuée à l'aide d'opérateurs de morphologie mathématique. Ensuite une méthode de contour actif a été employée afin d'extraire la frontière de ces structures. Cette méthode est fondée sur l'utilisation de descripteurs de Fourier, ce qui permet d'adapter simplement le modèle à la morphologie de chacun des objets du visage. Les contours, ainsi obtenus sur le visage à identifier, sont ensuite mis en correspondance avec ceux obtenus sur les visages d'une base d'apprentissage.

L'obtention des éléments du visage est effectuée par des algorithmes utilisés pour détecter les contours des éléments du visage qui sont de type pyramidal. Une courbe initiale entourant l'objet étudié composée au départ de deux demi-ellipses, puis la recherche d'une courbe  $v_1$  décrite par une seule harmonique. Le résultat obtenu est alors utilisé comme initialisation d'un modèle  $v_2$ , à deux harmoniques, jusqu'à l'obtention d'une courbe  $v_n$ , à  $N$  harmoniques, décrivant les contours de l'élément recherché. Le nombre  $N$  est fixé selon les résultats expérimentaux pour chaque type de contour. La figure 1.5 est une illustration pour le cas de la recherche des yeux et de la bouche.

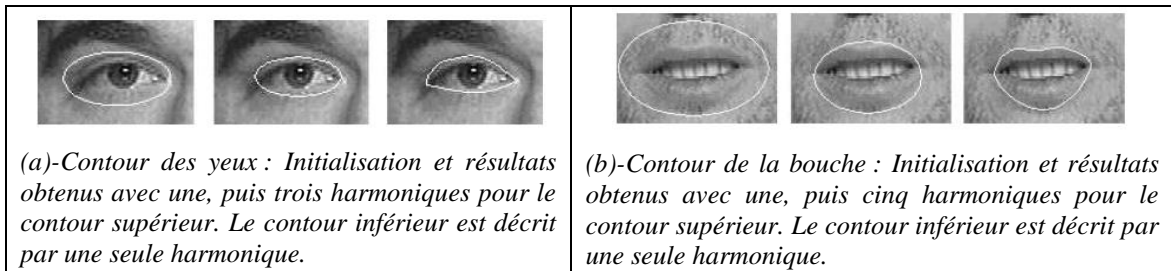


Figure I.5 : Extraction du contour des yeux / bouche.

Une autre approche de la méthode géométrique se base généralement sur la mesure des traits distinctifs du visage, de leur position dans l’image et de leurs positions les uns par rapport aux autres. La position et la largeur du nez, la position et la largeur de la bouche, l’épaisseur des sourcils et leurs positions verticales, comme le montre la figure 1.6, sont des mesures parmi les mesures importantes utilisées pour l’authentification d’un visage.

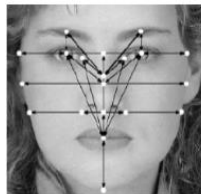


Figure I.6 : Visage accompagné de mesures géométriques (image extraite de [Cha 09])

Une phase d’enrôlement, pendant laquelle sont stockées les mesures prises sur les images de chaque personne parmi plusieurs personnes traitées, facilite l’identification du visage après comparaison de toutes ces mesures avec celles prises sur l’image de test pendant la phase d’identification. La classification et la décision sont généralement réalisées selon le degré de ressemblance maximale entre les deux séries de mesures. En effet, deux visages se ressemblent si la distance calculée entre les deux ensembles de mesures est minimale. Bien évidemment, si toutes les distances calculées sont supérieures à un seuil, le système considère que la personne est inconnue.

Il existe plusieurs techniques de prises de mesures, mais leur point commun est qu’elles se concentrent toutes sur la mesure de trois organes importants, à savoir : les yeux, le nez et la bouche.

Parmi ces techniques, celle qui permet d'extraire par exemple un vecteur de 8 mesures, correspondant notamment à des distances entre les organes [Cha 09] :

$$V = [D_{\text{Yeux}}, D_{\text{Nez}}, D_{\text{Yeux-Nez}}, D_{\text{Bouche}}, D_{\text{Bouche-Nez}}, \dots]. \quad (25)$$

D'autres techniques utilisent des mesures plus nombreuses afin de constituer la signature d'un visage. A titre d'exemple, le système WHO\_IS (système d'identification des visages humains) [Las 02] qui est caractérisé par le fait qu'il ne traite que la vue de face du visage, il est invariant à la présence des parties ou élément perturbateurs à la reconnaissance tel que la moustache, la barbes et indifférent aux expressions faciales, invariant également au port de chapeau, de casquette, de foulard, de lunettes. Le système WHO\_IS, est basé sur celui proposé par Kamel et al. [Kam 93] où ils avancent que 9 points bien choisis sont suffisants pour identifier un visage. Mais pour des bases de données plus larges le nombre de points caractéristiques doit être plus grand. Ainsi Le système WHO\_IS a réutilisé les 9 points du modèle de Kamel et al. [Kam 93] [Las 02] plus 3 points qu'ils ont rajoutés. Il s'agit, tel qu'illustrée par la Figure 1.7.

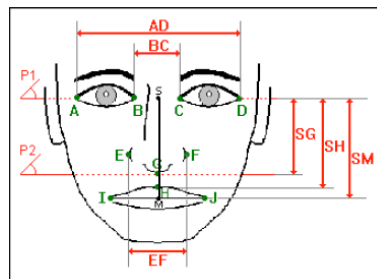


Figure 1.7: Modèle géométrique du visage dans le système WHO\_IS [Las 02].

Plus récemment, [Cha 09] a confirmé que la méthode géométrique associée à une classification par réseau de neurones donne 84.2% de taux de reconnaissance sur une base de 47 personnes.

### 1.6.2. Phase de Classification

La classification est l'attribution ou non d'un visage donné à une classe spécifique. La classe, ici, est une personne avec des images de visage dans la base de données et la

non-classe correspond au cas où le visage donné n'appartient pas à aucune classe parmi les différentes classes de la base de données. Cette répartition nécessite l'introduction d'une mesure de similarité.

### **1.6.2.1. Types d'apprentissage**

On peut classer les méthodes d'apprentissage en plusieurs catégories dont les plus distinguées sont l'apprentissage supervisé et l'apprentissage non-supervisé. Nous allons nous positionner dans le cas d'une tâche de classification. C'est ce dont nous avons besoin dans le cadre de cette thèse pour réaliser la reconnaissance de visage. Le but est d'associer à un objet donné une étiquette correspondant sa classe.

**Apprentissage supervisé** : Il a pour but d'apprendre par l'exemple. Etant données une liste d'objets avec leurs étiquettes de classe - appelée ensemble d'apprentissage – qu'il faut fournir au préalable afin que le système soit capable d'expliquer et ensuite de prédire l'appartenance d'un nouvel objet à une classe connue a priori. Beaucoup d'algorithmes d'apprentissage supervisé sont utilisés pour faire de la reconnaissance d'objets : caractères, visages, personnes...

La méthode la plus classique est certainement l'algorithme des «K-Plus Proches Voisins» (KPPV) [Ley 10], dont le principe consiste à calculer la distance d'un nouvel objet par rapport à ceux dont la classe est déjà connue, par exemple avec une distance euclidienne ; le nouvel objet appartiendra à la classe dont il est le plus proche. Cette technique donne d'assez bons résultats dans des cas simples et est facile à mettre en œuvre car elle est non-paramétrique.

Toutefois le temps de calcul de la prédiction est assez long car il nécessite un calcul de distance par rapport à tous les éléments de la base d'apprentissage. La figure I.8 montre un exemple simple d'une classification par un algorithme des K-plus proches voisins avec un calcul de distance euclidienne : le nouvel exemple est associé à la classe B car la majorité de ses plus proches voisins appartiennent à la classe B (au sens de la norme euclidienne).



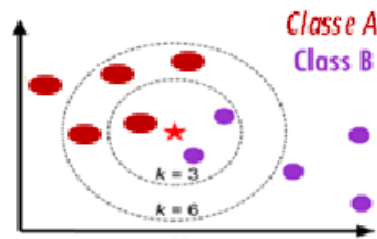


Figure 1.8 – Exemple simple de classification par un algorithme K-Plus Proches Voisins.

Les distances entre le nouvel exemple et les différents points des deux classes sont calculées. Comme ses plus proches voisins appartiennent à la classe B et A respectivement pour  $k=3$  et  $k=6$  il est alors étiqueté comme appartenant à cette classe.

**Apprentissage non-supervisé :** Pour ce type d'apprentissage, l'ensemble d'apprentissage est seulement composé d'exemples, sans aucune étiquette de classe. C'est à l'algorithme de trouver des dépendances, des structures entre les différents exemples. Le «clustering» ou partitionnement de données regroupe un ensemble de méthodes d'apprentissage non-supervisé, comme l'algorithme des K-means [Ley 10]. Les classes (ou «clusters» en anglais) sont créées par l'algorithme qui regroupe dans une même classe des objets ayant des caractéristiques communes entre elles et différentes avec les objets n'appartenant pas aux mêmes classes.

Comme exemple d'illustration, les K-means. L'algorithme est initialisé aléatoirement avec un certain nombre de clusters pour lesquels un point moyen, appelé centroïde, est déterminé. D'une itération à une autre, la distance entre chaque exemple aux différents centroïdes est calculée ; chaque exemple est alors associé au cluster dont la distance au centroïde est la plus proche. Puis les centroïdes sont réévalués. L'algorithme se termine lorsqu'il n'y a plus aucun changement.

### 1.6.2.2. Choix de la Base d'apprentissage

Pour réaliser un apprentissage, une base de données, appelée base d'apprentissage, est disponible ; elle est composée d'un grand nombre d'exemples. Le choix de la base d'apprentissage n'est pas trivial. Il faut tout d'abord choisir des images en rapport avec

l'application visée. Un autre point important est le nombre d'images prises dans la base d'apprentissage.

Chaque exemple de cette base d'apprentissage est caractérisé par un descripteur. Puis le classifieur est entraîné sur ces données. Son rôle est de déterminer les caractéristiques communes aux exemples d'une même classe afin de pouvoir ultérieurement reconnaître à quelle classe appartient un nouvel exemple inconnu. Parmi les points importants, pour assurer un apprentissage valide, le nombre de caractéristiques représentant un objet qui peut entraîner à une erreur de classification s'il n'est pas suffisamment décrit. Parmi toutes les méthodes existantes, on décrit donc ce qui suit l'apprentissage par réseau de neurones qui est inspiré du fonctionnement des neurones biologiques.

### **1.6.2.3. Technique des réseaux de neurones**

La technique des réseaux de neurones est une technique non-linéaire d'extraction de signatures utilisée pour la reconnaissance de visages [Cha 03], elle repose sur un réseau de neurones artificiels (RNA). Elle a été inspirée de la physiologie du système nerveux si parfaitement créée et conçue. Le neurone formel, introduit par J.Mc Culloch et W. Pitts dans les années quarante, constitue la base de l'architecture des RNA.

En se basant sur la théorie de l'analogie, le modèle de neurone formel a été introduit pour tenter de reproduire le fonctionnement des cellules neuronales. Dans le système nerveux biologique, on considère que le neurone est l'élément de base assurant à la fois la propagation et le traitement de l'information. Chaque neurone reçoit, au niveau des dendrites attachées à son corps cellulaire via des synapses ou liaisons inter-neurones, des stimulations électriques ou chimiques venant d'autres neurones. Lorsque le neurone a reçu par ce biais un nombre suffisant de stimulations, il transmet à son tour une information vers des neurones voisins, grâce à un influx nerveux qui se propage le long de son axone (cf. figure I.9-(a)).

Un neurone formel est une fonction algébrique non-linéaire (paramétrée et à valeurs bornées) de variables réelles appelées entrées, qui tente de reproduire ce mode de fonctionnement. Un neurone formel effectue une somme pondérée des signaux d'entrée qui lui parviennent. Cette somme pondérée sert de paramètre à une fonction, souvent non-linéaire, qui la transforme en nouveau signal transmis à la sortie (cf. figure I.9-(b)).

En référence au fonctionnement du neurone biologique, la fonction intervenant après la sommation des entrées est appelée fonction d'activation.

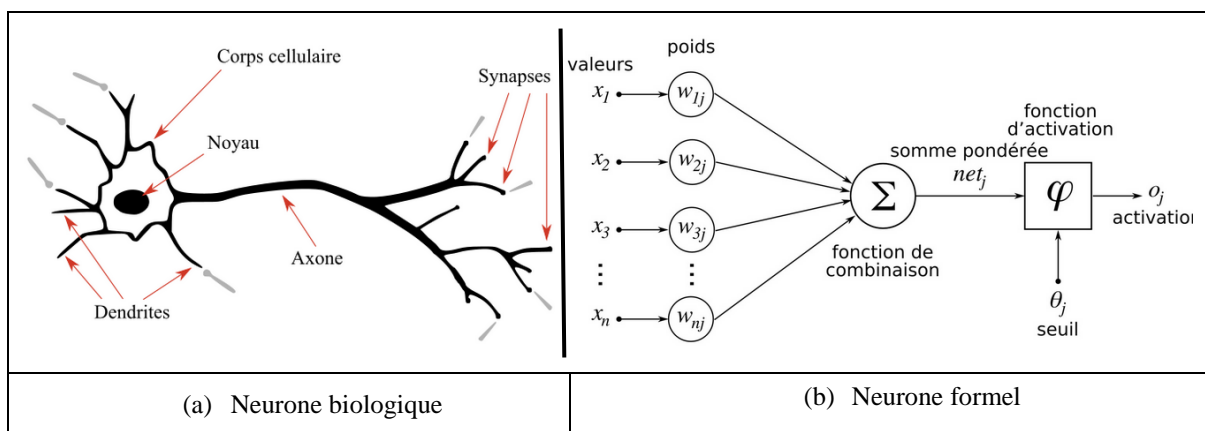


Figure I.9 : Analogie entre le neurone biologique et le neurone formel

Le neurone formel est une unité élémentaire d'un réseau neuronal. Il effectue la somme pondérée de ses entrées, car la valeur de cette sommation est comparée à un seuil. La sortie du neurone est la suivante:

$$\mu = \sum_{i=1}^n w_i \cdot x_i - \theta \tag{26}$$

$$\varphi = f(\mu) \tag{27}$$

Tels que:

$x_i$  : Élément du vecteur d'entrée,

$i=1\dots n$  : Entrées neuronales formelles,

$Y$  : Sortie,

$w_i$  : Paramètres de pondération,

$\varphi$  : La fonction d'activation.

La façon par laquelle un réseau de neurones (RNA) peut servir à extraire la signature associée à un visage, un apprentissage permet d'adapter les paramètres du réseau durant la phase d'enrôlement, afin qu'il puisse par la suite discriminer les signatures présentées sur ses entrées. Pour ce faire, le processus est réalisé de façon itérative. Un visage de la base d'apprentissage sélectionné dont sera calculée la signature qui sera placé en entrée du réseau. Les sorties du réseau pour cette valeur de ses entrées sont calculées, puis l'évaluation de la fonction d'erreur en considérant que les sorties idéales sont toutes nulles sauf celle qui correspond à la classe dont fait partie l'individu. Pour ajuster les paramètres du réseau, la technique standard de rétro-propagation de l'erreur est ensuite utilisée. Cette procédure est itérée en présentant à l'entrée, plusieurs fois et dans un ordre aléatoire, toutes les signatures extraites des images de la base.

La signature du visage cible Durant la phase d'identification est calculée et appliquée en entrée du RNA discriminant. La sortie du réseau qui prend la valeur maximale indique à quelle classe appartient le visage cible. L'introduction d'un seuil sur la valeur de la sortie maximale permet de déterminer si l'image cible ne correspond à aucun individu connu de la base.

## **Conclusion**

Dans ce premier chapitre, nous avons développé les principales techniques de reconnaissance de visages. Nous avons, tout d'abord, abordé une description détaillée de la biométrie en précisant que la reconnaissance de visage est l'une des modalités biométriques importante pour identifier une personne. Ensuite nous avons décomposé un système biométrique en trois modules caractérisant les différentes phases de reconnaissance faciale tout en indiquant les critères d'évaluation de ce système.

Par ailleurs, nous avons classifié les techniques de reconnaissance de visages en trois grandes phases, à savoir la phase de détection de visage, celle d'extraction des caractéristiques de ce visage et la phase de reconnaissance de ce visage. Nous avons distingué différents types de méthodes de localisation faciale telles que les approches

avec ou sans segmentation qui se distinguent par le fait que l'analyse s'effectue par une segmentation explicite du visage pour le premier cas ou par un traitement particulier pour certaines composantes du visage, la méthode de Viola et Jones en décrivant les principaux éléments de cette méthode, la méthode basée sur l'analyse de la couleur de la peau considérée efficace pour la détection de visage dans une image, et enfin la méthode Template Matching consistant à manipuler des sous régions de l'image pour détecter l'endroit le plus proche de la présence de visage dans l'image.

Toutefois, nous avons soulevé la difficulté de l'ACP, à gérer les occultations et les variations de poses, d'illumination et l'expression. Et nous avons, également, présenté deux grandes méthodes, pour la phase d'extraction des caractéristiques du visage, les moments de Hu et les moments de Legendre qui ont été utilisés dans de nombreuses applications de reconnaissance de formes malgré leurs résultats qui restent un peu pertinents en les comparant avec les méthodes géométriques qui utilisent de nombreuses mesures afin de construire la signature du visage à identifier.

En fin, pour la phase de reconnaissance de visage, nous avons entamé la méthode des K-plus proches voisins et celle des réseaux de neurones qui prennent en considération le choix de la base des données et le nombre d'images dans celle-ci.

Dans le cadre de cette thèse, nous proposons, dans le chapitre suivant, deux approches pour la mise en situation de la phase de détection de visage, en effet, nous commencerons par une approche de détection de visage basée sur les histogrammes qu'on a constaté moins efficaces par rapport à la deuxième approche basée sur la couleur de la peau pour détecter un visage dans un document mixte.

## 2. Contribution à la détection de visage dans un document mixte

### 2.1. Introduction

Après avoir présenté le contexte scientifique et l'état de l'art, nous décrivons dans ce chapitre les aspects théoriques et pratiques de notre contribution sur la méthode proposée pour la détection de visages. Nous proposerons d'abord une première approche, basée sur les histogrammes, pour la phase de détection de visage, qu'on a constaté pas assez performante et nous proposons, par la suite, une amélioration par une autre approche utilisant les méthodes Hybride et Template Matching.

En effet une mise en évidence des limitations de la première approche en termes de non généralisation, nous a ramené à proposer une méthode qui comble ces lacunes. Nous décrivons également les modules composant le système de reconnaissance de visages, dans sa phase de détection, exploitant notre nouvelle méthode de détection de visage dans un document mixte composé de texte et d'images. Nous détaillons les traitements réalisés dans chaque module, suivant l'architecture standard présentée dans le chapitre 1 et rappelée par le synoptique de la figure 2.1.

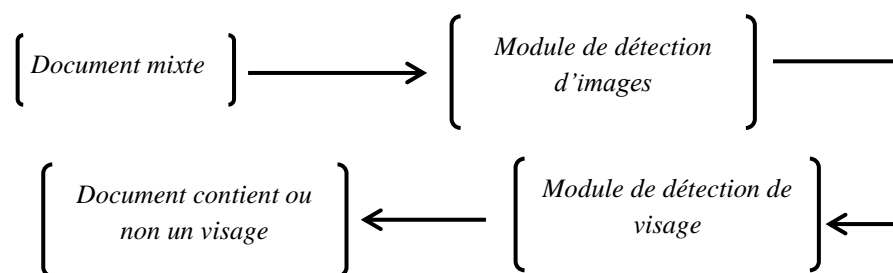


Figure 2.1. Synoptique de notre système dans sa phase de détection de visage

Ce chapitre comporte quatre sections principales. Dans la première, nous présentons les aspects théoriques qui nous ont motivés à explorer cette piste. Nous exposons ensuite un aperçu général du système que nous avons conçu pour la reconnaissance de visages. Puis nous détaillons l'aspect pratique du système concernant la première vision. La deuxième partie explique les lacunes rencontrées dans l'approche précédente pour passer à développer l'aspect pratique de la deuxième approche. Et en troisième section

nous donnons une définition de l'objectif et énumération de ses conséquences puis on termine ce chapitre par les résultats et discussions concernant la seconde approche.

## 2.2. Approche des histogrammes

### 2.2.1. Notion d'histogramme

Pour déterminer un histogramme on a à compter le nombre de pixel pour chaque intensité de l'image. La quantification permet d'approcher un signal continu par les valeurs d'un ensemble discret d'assez petite taille c.à.d. qui regroupe plusieurs valeurs d'intensité en une seule classe, ce qui peut permettre de mieux visualiser la distribution des intensités de l'image.

Les histogrammes sont considérés comme des outils fondamentaux utilisés dans de nombreuses applications. Ils sont fréquemment utilisés dans le traitement d'images. Cependant les histogrammes multidimensionnels ont l'inconvénient d'occuper en mémoire un volume gigantesque [Tou 05]. Ce volume est de quelques dizaines ou centaines de mégaoctets pour les images couleur à trois composantes.

Partant du fait qu'une image couleur (à titre d'exemple figure 2.2) peut être décomposée en trois images uni-modales indépendantes (figure 2.3), certaines méthodes de traitement d'images utilisent les histogrammes mono-dimensionnels R, V et B de ces trois images comme c'est illustrée dans la figure 2.4.



Figure 2.2 : Image Lena en codage RVB



Figure 2.3 : Composante R(a), composante V(b), composante B(c)

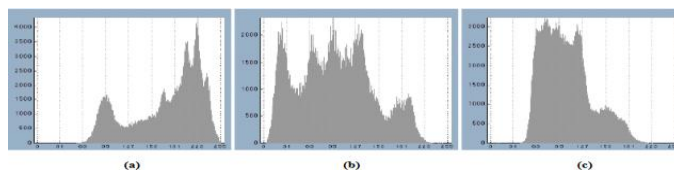


Figure 2.4 : Histogramme R (a), histogramme V (b), histogramme B(c)

## 2.2.2. Description de l'approche

Après la recherche dans un ensemble de travaux de la littérature, qui s'intéressent au domaine de la reconnaissance de visage, nous avons proposé la première version du système destiné à la détection du visage dans un document mixte, ce module de détection se base sur un ensemble de traitements qui seront détaillés dans ci-dessous. L'architecture du système développé est illustrée dans la figure 2.5.

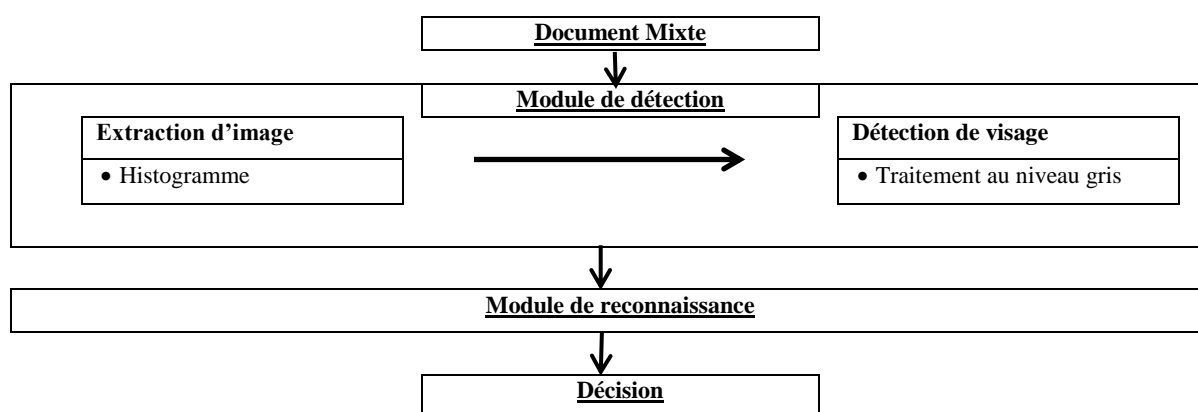


Figure 2.5. Première vision décrivant la phase détection du système envisagé à développer

La première version du système élaboré, au niveau de la phase de détection, est composée des phases suivantes :

### Phase 1 : Sélection d'un document contenant un visage

La détection d'un visage, dans un document mixte, n'est aussi facile sans hypothèse. C'est la raison pour laquelle notre système va prendre en considération l'hypothèse d'un document texte contenant une image simple (sans la présence, dans l'image, d'autres figures qui peuvent perturber les résultats du système). On a donc établi une base de documents mixtes qui sont composés, à la fois du texte et une image contenant un visage (figure 2.6), afin de pouvoir sélectionner, à chaque fois, un de ces documents qui sera l'entrée du système. A partir de cette entrée, le système doit extraire la zone du document contenant le visage, après détecter le visage qui, par la suite, sera comparé aux visages de la base de données afin d'identifier la personne dans l'image d'entrée.





Figure 2.6. Différents exemples de documents, utilisés, contenant un visage

**Phase 2 : Extraction de l'image, contenant le visage, du document.**

Cette phase a été réalisée en deux étapes :

- *Etape 1 :*

L’histogramme vertical du document nous a permis de déterminer l’intervalle où l’intensité des pixels est plus élevée, d’où on peut en déduire les références du premier pixel  $x_1$  (la borne haute) et du dernier pixel  $x_2$  (la borne basse) de l’image dans le document, comme c’est illustré dans le tableau 2.1, afin d’établir la première extraction du document initial en effet on a pu extraire, du document, la partie illimité par ces deux points  $x_1$  et  $x_2$ .

- *Etape 2 :*

Suite à l’étape précédente, et en se basant sur la partie extraite dans cette étape 1, l’histogramme horizontal nous a permis de déterminer, de même, l’intervalle où

l'intensité des pixels est assez élevée d'où on peut repérer les références du premier pixel  $x_3$  (la borne gauche) et du dernier pixel  $x_4$  (la borne droite) de l'image dans le document comme c'est illustré da tableau 2.1 afin d'établir la deuxième extraction , en effet on a pu extraire, du document, la partie illimité par ces deux points  $x_3$  et  $x_4$ . Ces deux étapes sont illustrées par le tableau 2.1 présenté ci-dessous.

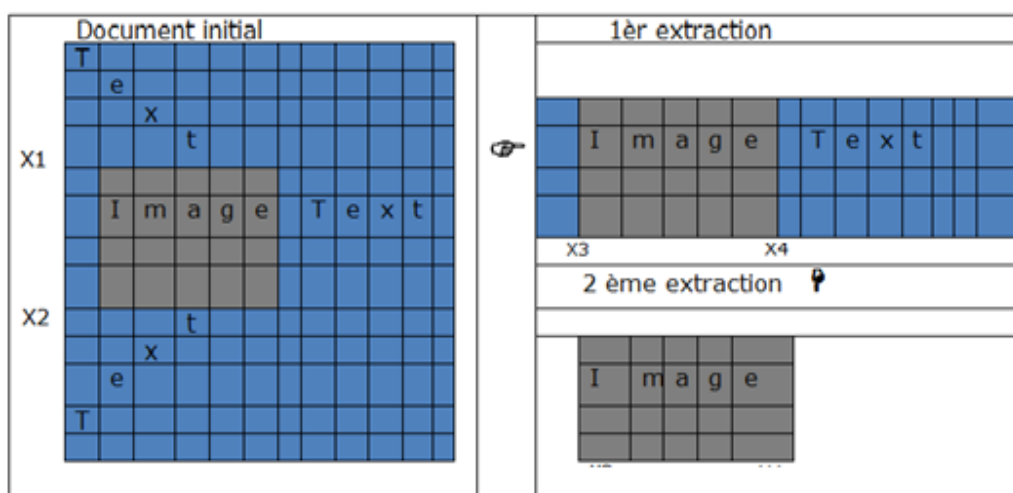


Tableau 2.1 : Etapes d'Extraction de l'image du document

Cette phase nous a permis, donc, comme le montre la figure 2.7, de délimiter la zone image contenant le visage, à partir du document initial pour l'exploiter dans les phases suivantes.

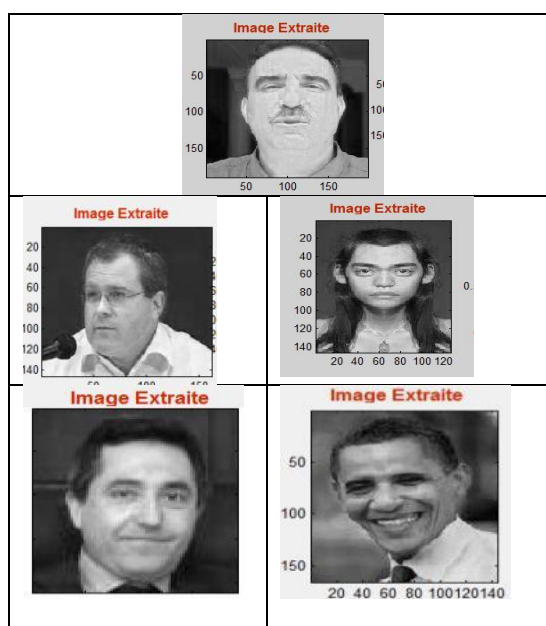


Figure 2.7. Extraction de l'image

**Phase 3 : Détection du visage dans l'image.**

Cette phase commence par un prétraitement sur l'image extraite dans la phase précédente contenant le visage à détecter afin de déterminer le contour de ce visage suivi d'un traitement permettant de localiser la partie visage de la partie non-visage afin d'extraire uniquement le visage sur lequel le système va se baser dans la phase suivante. Ce processus est illustré par la figure 2.8.

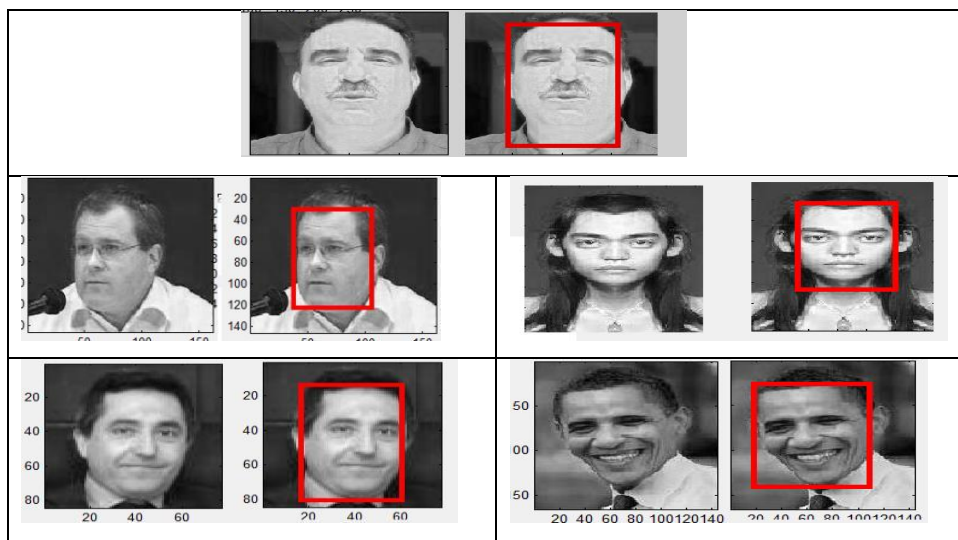


Figure 2.8. Détection du visage

Dans cette première partie, on a utilisé l'histogramme vertical et l'histogramme horizontal dont le but de déterminer la zone, du document teste, qui présente un niveau de gris élevé qui correspond à l'image présenté dans le document. Un traitement à base de la détermination du contour de cette image nous a ramené à encadrer le visage se trouvant dans cette image pour l'extraire et ensuite extraire ses caractéristiques.

Nous proposons, dans la suite, d'exploiter et de développer la méthode utilisée par ce système visant à tenir en compte d'autres hypothèses visant, aussi bien, le type de texte du document et l'image contenant le visage à extraire du document. Cependant, il faut instaurer des règles permettant la détection des visages tenant compte de minimiser le temps d'exécution et d'exploiter une base des données importante aussi bien au niveau nombre de classes d'individus qu'au niveau nombre de visages pour chaque classe tout en maximisant le taux de détection.

## 2.3. Approche basée sur les méthodes ‘Hybride’ et ‘Template Matching’

En ce qui concerne la deuxième approche de notre recherche, nous présentons, un système de reconnaissance de visages dans un document mixte (texte + images) basé, dans sa première phase de détection de visage, sur la méthode Hybride permettant de distinguer, à partir du document mixte, deux catégories d’objets, la première catégorie est tout ce qui est textes ou images contenant des images ne présentant pas la couleur de peau et la deuxième catégorie est toute image ayant la même couleur que la peau, puis en deuxième phase de détection de visages, basé sur la méthode Template Matching permettant de distinguer parmi les images de la deuxième catégorie uniquement celles qui contiennent des visages pour pouvoir les détecter (figure 2.9).

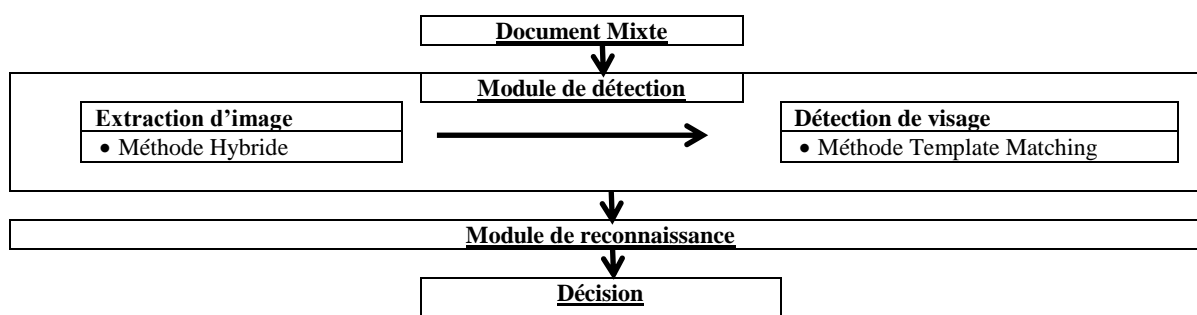


Figure 2.9 : Description de la phase de détection du système développé dans la deuxième approche

Etant donné que l’étape détection de visage est une étape primordiale, on a pensé, comme deuxième approche, à améliorer cette étape, par rapport à la même étape concernant la première approche. Cette amélioration consiste à remplacer la technique des histogrammes par deux méthodes Hybride et Template Matching :

-La méthode Hybride : qui permet une réduction considérable de l’espace de recherche des visages qui va se limiter uniquement aux sections de couleur clair comme on l’a détaillé dans le chapitre 1. Les performances du système de détection des visages sont donc étroitement liées avec les résultats obtenus lors de la segmentation de la couleur de peau. Nous présentons, dans ce travail, une étude comparative sur la détection de la couleur de peau par seuillages selon les espaces chromatiques RGB. L’étape de seuillage nous fournira des images binaires dont les sections peau prendront le plus

haut niveau de gris soit 1 (blanc) alors que les autres sections (non peau) prendront le plus bas niveau de gris soit 0 (noir).

-La méthode Template Matching dont le principe a été détaillée, également, dans le chapitre 1 et à titre de rappel en voilà son idée général : il s'agit, étant donné une image quelconque et un modèle prédéfini, de déterminer si ce modèle figure dans l'image à tester et d'en trouver la position.

Lors de la phase de détection, notre système est testé, avec une variété de documents contenant plusieurs images dont figurent, ou non, des visages de personnes sans conditions sur l'éclairage et sans conditions sur l'angle de rotation du visage dans l'image.

## 2.4. Objectif de la deuxième Approche

L'objectif de la seconde approche est de ne pas se limiter sur un document où figure une seule image, mais de généraliser cette situation pour permettre à l'utilisateur du système développé de détecter l'ensemble des visages des personnes présentés dans un document mixte. Un scénario type de l'utilisation de ce système se déroule en trois phases en utilisant deux méthodes.

### **Phase 1 : segmentation des régions ayant la couleur de la peau.**

En se basant sur la méthode Hybride (basée sur la couleur de Peau), développée dans le chapitre précédent, cette phase consiste à sélectionner des pixels représentant la couleur de la peau à partir d'un document mixte d'entrée puis filtrer les zones où les pixels, en masse, sont de la même couleur que celles de la couleur de la peau pour masquer les zones non couleur de la peau.

### **Phase 2 : segmentation des régions contenant des visages**

En se basant sur la méthode Template Matching, développée aussi dans le chapitre précédent, cette phase s'intéresse à éliminer, à partir des résultats de la phase précédente, les régions ayant la couleur de peau et qui ne présentent pas un visage pour garder uniquement les régions ayant la couleur de la peau,

### **Phase 3 : Détection des visages**

Dans cette phase la détection des visages est établie par appariement du gabarit « Template Matching » qui s'intéresse à comparer l'intensité des pixels entre un gabarit prédéfini et plusieurs sous-régions de l'image à analyser. Ce processus consiste en pratique à effectuer plusieurs balayages couvrant toute la superficie de l'image. Les endroits les plus propices à la présence de visages seront donc facilement identifiés par des minimums de distance entre le gabarit et l'image sous-jacente. La figure 2.10 illustre le processus détaillé des différentes phases indiquées ci-dessous :

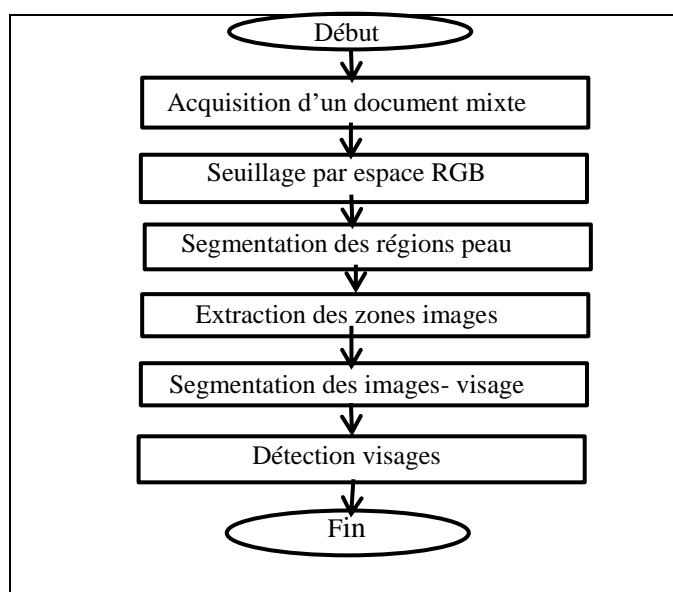


Figure 2.10. Processus reflétant l'objectif général du système développé

### 2.4.1. Détection des sections ayant la couleur de la peau

Plusieurs modèles de couleurs pouvant s'appliquer à la détection de la peau cette variabilité dépend de l'espace de couleur adopté pour la représentation des pixels. Les modèles les plus utilisés sont : l'espace RGB, RGB normalisé, HSV, YCbCr. Dans ce travail nous avons choisi l'espace RGB qui traduit la physiologie de l'œil. En effet l'œil humain distingue les couleurs à travers des récepteurs appelés cônes. Il existe en tout et pour tout trois types de cônes, ce qui signifie que toute couleur discernable par l'homme est caractérisée par un point dans un espace à trois dimensions. Plus précisément, les valeurs, mesurées, par les trois types de cônes représentent essentiellement les intensités du rouge, du vert et du bleu,

l'espace RVB étant défini à partir de ces trois couleurs primaires. Son codage a été mis au point en 1931 par la Commission Internationale de l'Eclairage (CIE) [Ben 07].

La peau est représentée, dans la majorité des cas, par une portion d'un espace de couleurs particulier. Pour isoler les régions susceptibles de contenir des visages, en utilisant les frontières de ces régions comme valeurs de seuillage sur une image, il est possible d'extraire les pixels dont la couleur peut s'apparenter à celle de la peau [Gar 99]. En effet, pour détecter les régions de peau, un seuil  $S$  caractérisant la couleur de la peau étant défini. Ainsi pour détecter les pixels couleur peau, la règle est :

$$\begin{cases} 1 & \text{si } x < S \\ 0 & \text{si } x \geq S \end{cases} \quad (28)$$

La figure 2.11 présente un exemple établi par notre système qui reflète la conversion d'un document mixte couleur (texte + images) en deux catégories de pixels soit les pixels blancs qui représentent la couleur de la peau et les pixels noirs qui représentent la couleur non peau.

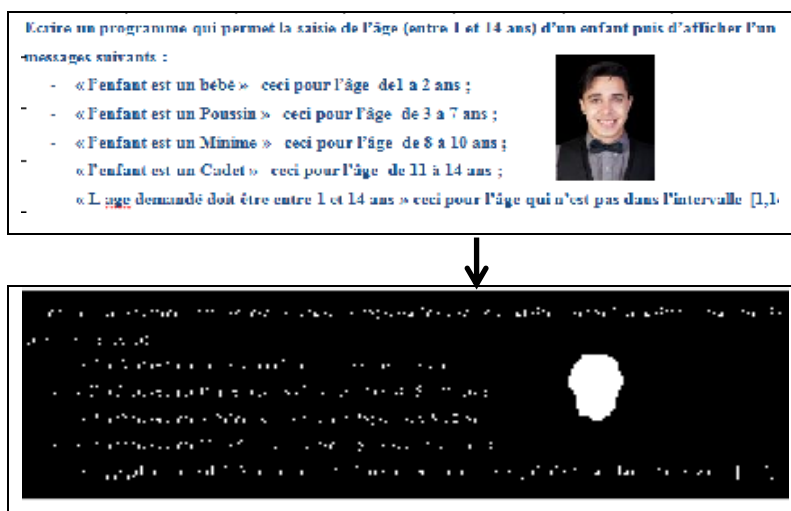


Figure 2.11. Exemple de document mixte converti en binaire développé par notre système.  
 Pixels blancs = couleur de peau, pixels noirs = couleur non peau.

## 2.4.2. : Filtrage des régions ayant la couleur de la peau

Lors de cette phase, un traitement est appliqué à l'ensemble des pixels du document résultat de la phase précédente qui consiste à parcourir (pixel par pixel) tous les pixels et détecter les pixels  $x$  tels que :

$$\begin{cases} 1 & \text{si } x = 1 \text{ et } x + 1 = 1 \text{ et } x + 2 = 1 \\ 0 & \text{sinon} \end{cases} \quad (29)$$

Ainsi les pixels ne vérifiant pas la règle (19) sont forcés de prendre la couleur noir malgré leur couleur initial blanc, par conséquent seuls les pixels, consécutifs, blanc qui figurent dans une image couleur peau qui prennent la couleur blanche. La figure 2.12 présentée ci-dessous illustre un exemple appliqué à l'exemple présenté dans la figure précédente.



Figure 2.12. Exemple de document filtré.

### 2.4.3. Extraction des régions des images couleur peau et détection de visage

Au cours de cette étape, un balayage permettant de localiser les zones qui présentent des pixels blanc. La difficulté de cette phase réside dans le fait que le document acquis en entrée peut contenir plusieurs images couleur de peau (figure 2.13). Dans ce cas, le processus consiste à détecter la première image dont les pixels sont blancs, lui subir un traitement impliquant la méthode Template Matching pour détecter si celle-ci contient ou non un visage, puis itérer le même traitement pour détecter la présence ou non de visage dans les autres zones dont les pixels sont blancs.

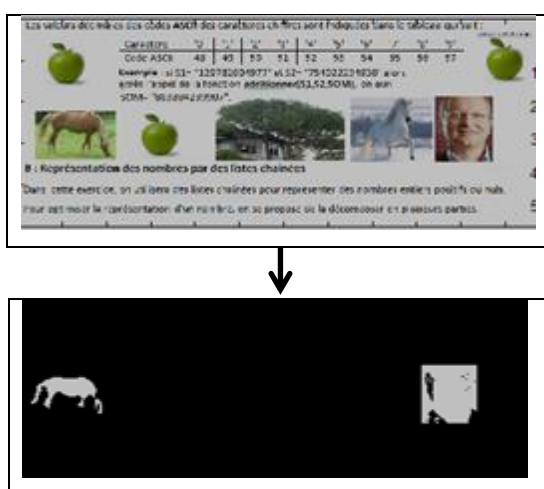


Figure 2.13. Exemple de localisation de deux images couleur peau dans un document mixte.



La méthode Template Matching, consiste à balayer chaque sous-région parmi les régions de l'image à analyser, et comparer l'intensité des pixels entre un gabarit model (Template) prédéfini (dans notre cas figure 2.14) et chaque sous-région de l'image à analyser. Ce processus effectue, en pratique, plusieurs balayages couvrant toute la superficie de la zone. Les endroits les plus propices à la présence de visages seront donc facilement identifiés par des minimums de distance entre le gabarit et la zone sous-jacente. La distance envisagée, dans notre cas, est la norme  $L_2$  indiquée dans la formule (19).

Nous présentons dans la Figure 2.14 le modèle qui a été utilisé par notre système pour détecter si une image contient ou non un visage. Ce modèle est une zone parmi les zones d'un visage avec laquelle on peut décider si une image contient ou non un visage, il s'agit de la zone qui encadre les yeux d'un visage.



Figure 2.14. Modèle de Template utilisée pour détecter si une image contient un visage.

En fixant un seuil pour la position du segment détecté supposé être plus similaire à la template, nous pouvons juger que, grâce à ce seuil, si ce segment contient ou non un visage. La figure 2.15 est un exemple qui illustre ce processus, en effet, la partie (a) de cette figure reflète la position du gabarit sur le visage par contre dans la partie (b), la position du gabarit est au-delà du seuil ce qui signifie qu'il n'y a pas de visage dans cette image considérée, au départ, une image ayant la couleur peau.

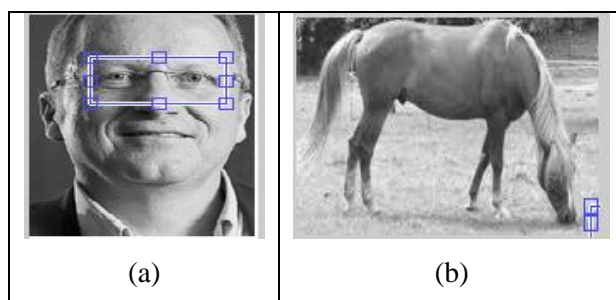


Figure 2.15. Détection de visage par méthode Template Matching (image pour chaque cas).

Les pixels de l'image qui ne contiennent pas de visage seront alors envoyés au niveau de gris le plus bas soit 0 (noir), par conséquent seuls les pixels de l'image contenant un visage (la figure 2.16) qui seront de niveau gris 1 (Blanc). Toutefois il a été appliqué un filtre pour éliminer les petites zones ayant la même couleur que celle de la peau car ces

zones, vues leur petite dimension, il est supposé qu'elles ne correspondent pas à un visage, d'où un rejet de ces zones entraînant que leurs pixels seront au niveau de gris 0.



Figure 2.16. Segmentation de l'image contenant un visage.

Pour détecter le visage, d'abord, nous devons extraire l'image binaire du document. Nous essayons ensuite de localiser la zone du visage, en fonction de la position du modèle identifié dans le paragraphe précédent comme repère initial. Le résultat est une image en niveau de gris représentant le visage à détecter en question comme le montre la figure 2.17.

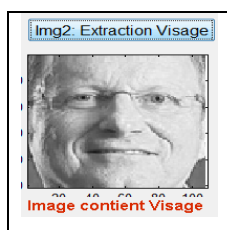


Figure 2.17. Extraction du visage

## 2.5. Résultats et discussions

### 2.5.1. Interface graphique du système développé dans sa phase de détection

L'interface développée pour notre système, comme le montre la figure 2.18 présentée ci-dessous, consiste à sélectionner un document où figure du texte et des images variés impliquant ou non des visages, ensuite extraire les zones d'images correspondant aux visages en appliquant le module de détection d'images ayant la couleur de la peau (méthode Hybride) et le module de détection des visages basé sur la méthode Template Matching.

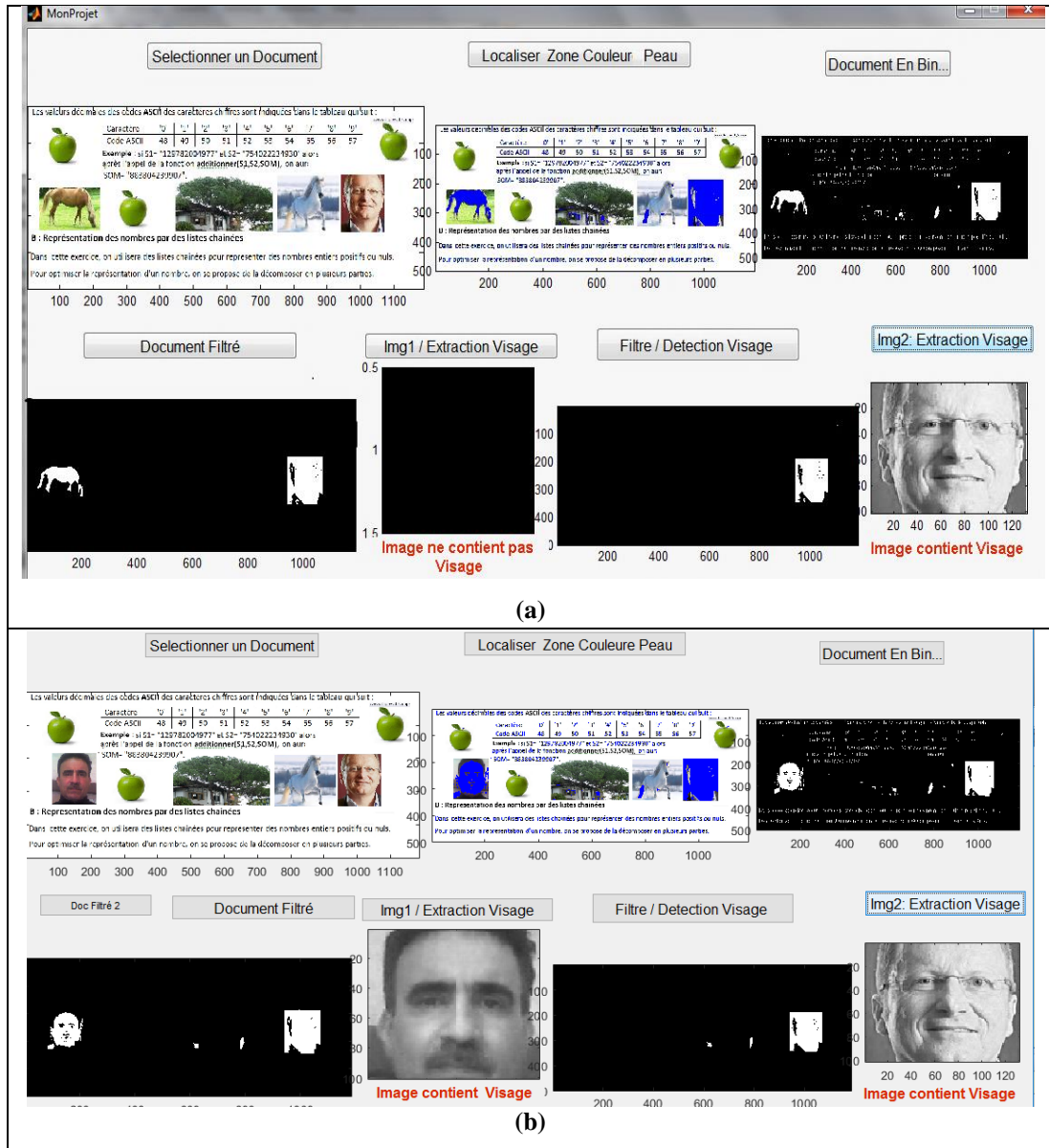


Figure 2.18. Interface du système développé  
 (a)- Cas d'un document contenant un seul visage ;  
 (b)- Cas d'un document contenant plus qu'un visage ;

### 2.5.2. Résultats de l'approche des histogrammes

On a utilisé une base de données contenant trente documents mixtes dont chacun des vingt-quatre parmi ces trente présente un visage tandis que chacun des six autres est composé de texte et d'image qui ne présente pas de visage. Les résultats expérimentaux, pour la détection de visage, basés sur la méthode des histogrammes, ont été réalisés ; aussi bien, sur les documents qui présentent un visage que sur les documents qui ne présentent pas de visage.

Les résultats expérimentaux ont montré que le taux de détection de visage dans un document mixte par la méthode des histogrammes est de 83%. Ces résultats sont légèrement meilleurs par rapports aux résultats souhaités. Parmi les paramètres qu'on a pris en considération, pour améliorer ce taux de détection, la résolution par axes d'histogramme. L'augmentation de la résolution donne une augmentation du taux de détection mais qui reste légèrement stable par rapport au taux signalé au paravent.

### **2.5.3. Résultats de l'approche basée sur la méthode Hybride et la méthode Template**

Dans les expérimentations de la section précédente, nous avons utilisé la méthode des histogrammes qui a fourni un taux de 83%. Les résultats de cette section indiquent que le taux de détection de visage peuvent être améliorés en ajoutant d'autre paramètres mais pour une base de donnée limitée. Ce qui nous a ramené à la réflexion de changement de méthode pour améliorer le taux de détection même pour une base de données plus importante. Ainsi nous avons pensé à une combinaison de la méthode Hybride et de la méthode Template Matching.

Nous avons effectué quelques tests sur des documents textes dont lesquelles on a intégré, aléatoirement, des images que nous avons téléchargé du Net. Les tests sont effectués sur des images de couleurs différentes et présentant des figures qui peuvent ou non être des visages. Nous notons dans la majorité des tests (85% des tests) des résultats positifs pour les deux types de documents qui présentent ou non des visages. Parmi les fausses détections de visage, des tests sur une image contenant un visage d'un enfant ou d'un bébé comme l'indique la figure 2.19. Cette fausse indication est due à l'écart de la similarité entre le gabarit (la Template Matching) et chaque fenêtre de l'image du document d'entrée lors de la segmentation à la recherche du segment de l'image le plus proche du gabarit.



Figure 2.19 : Cas d'un document ou figure un visage d'enfant non détecté

## Conclusion

Le système de reconnaissance, dans sa phase de détection, que nous avons développé est introduit dans cette thèse comme une étape préliminaire et de grande importance dans tout système de reconnaissance faciale, parce que avant de lancer le système pour identifier des individus dans un document mixte, on vérifie, d'abord, la présence ou non d'un visage dans le document qui est très bénéfique du fait qu'elle permettra au système de détourner l'opération de reconnaissance si la scène ne contient aucun visage.

Notre application repose, d'une part, sur la méthode hybride qui effectue la segmentation de deux différentes classes de zones, les premières zones du document, pouvant contenir un visage (à base de recherche la couleur de la peau) et les secondes zones dont la couleur diffère de celle de la peau. L'espace de représentation des couleurs sélectionné est l'espace RVB. Et d'autre part, l'application est basée sur la méthode d'appariement de modèles pour détecter la présence ou non, dans le document en entré, d'un visage.

Les résultats sont satisfaisants puisque ils nous ont permis d'augmenter les performances du système pour la détection du visage en comparaison avec la méthode des histogrammes. Cependant, pour mieux contourner le problème de la détection des visages, nous proposons comme perspectives à court terme d'essayer d'autres espaces colorimétriques tels que l'espace RVB normalisé et l'espace HSV et de comparer les résultats.

Dans le prochain chapitre, nous allons exposer notre contribution sur la phase de reconnaissance de visage, nous utiliserons deux approches différentes, la première approche basée sur les moments de Hu et de Legendre et la deuxième approche basée sur les méthodes géométriques et nous concluons par une au point de comparaison entre ces deux différentes approches.

## 3. Contribution à la reconnaissance de visage dans un document mixte

### 3.1. Introduction

Notre but étant de développer un système d'authentification de visage simple et efficace, il s'est avéré qu'il est nécessaire de travailler à tous les niveaux du système (détection, extraction des caractéristiques et reconnaissance) en apportant des contributions pertinentes à différents points de la chaîne de traitement. Ainsi, nous avons développé, par l'approche des histogrammes, une technique, pour la localisation de visage, qu'on a détaillé dans la première partie du deuxième chapitre et qu'on a amélioré par les méthodes Hybride et Template Matching, dont l'amélioration a été détaillée dans la seconde partie du même chapitre.

L'extraction de la région du visage à partir de l'image segmentée est une étape primordiale dans le processus de reconnaissance faciale. L'algorithme d'extraction doit être efficace et doit fournir, sans aucune ambiguïté, la région du visage dans l'image. Nous allons proposer une méthode, dans ce chapitre, comme première approche dans le cadre de cette thèse basée sur les moments invariants. Cette approche exploitera les moments de Hu et de Legendre. Et du fait qu'on a constaté que les résultats n'étaient pas assez pertinents, nous nous sommes ramenés à proposer de remplacer les algorithmes des moments invariants par une deuxième approche basée sur un algorithme exploitant une méthode géométrique efficace pour l'extraction des caractéristiques du visage détecté dans le document à analyser.

Une autre contribution de cette thèse consiste à développer une technique, basée sur les méthodes géométriques, impliquant, au niveau classification, l'algorithme des k-plus proche voisin et des réseaux de neurones qui prend en entrée les différentes caractéristiques du visage déterminées dans la phase d'extraction au nombre de dix-sept au lieu de sept(cas des moments de Hu) et dix(cas des moments de Legendre).

Les différents tests réalisés sur la base des données Feret-color ont permis de valider cette approche. Enfin, nous avons décidé de fusionner, pour la chaîne de traitement (détection + extraction + classification + reconnaissance), les approches les plus performantes présentées dans le cadre de cette thèse.

## 3.2. Reconnaissance de visage : Première Approche proposée

Etant donné que la chaîne de traitement, au niveau reconnaissance, passe, en premier lieu, par l'extraction de la région du visage à partir de l'image segmentée, qui est une étape primordiale dans le processus de reconnaissance faciale. Nous présenterons d'abord l'algorithme d'extraction développé comme première approche (figure 3.1) dont les résultats ne sont pas très satisfaisants. Et comme cette étape devrait être efficace et doit fournir, sans aucune ambiguïté, la région du visage dans l'image, nous allons proposer une autre technique améliorée dans la section suivante.

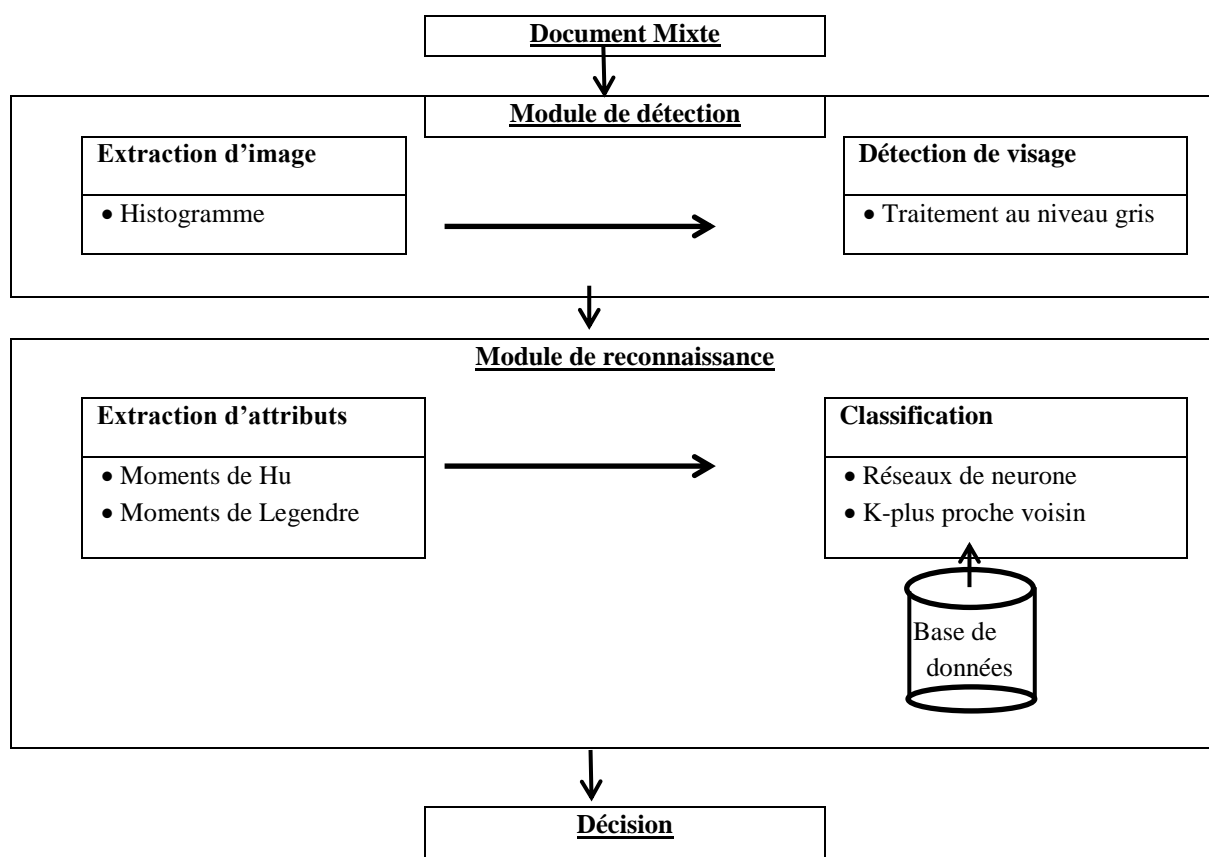


Figure 3.1 : Système élaboré en utilisant la première approche proposée.



### 3.2.1. Extraction des caractéristiques du visage

Nous avons utilisé, dans cette phase, les moments de Hu et les moments de Legendre (détaillés dans le chapitre 2) afin d'obtenir, d'une part, les caractéristiques du visage détecté dans la phase précédente, et d'autre part de créer, pour chacun des moments, une base de référence, pour l'ensemble des images de la base des données. Chacune de ces bases de références sera exploitée dans la phase suivante selon le choix du classificateur à utiliser dans cette phase.

### 3.2.2. Classification

#### 3.2.2.1. Réseau de neurones

La classification, par réseau de neurones, nécessite d'abord une première étape d'apprentissage dans laquelle sera défini la construction de la structure du réseau de neurones multicouches (le nombre de neurones de la couche d'entrée du réseau dépend de la taille du vecteur caractéristique, le nombre de neurones de la couche cachée et le nombre de neurones de la couche de sortie). Pour notre cas, il s'agit de 7 entrées indiquant les caractéristiques du moments de Hu, 10 cellules de la couche cachée et 20 sorties correspondant aux nombre de classes de visages dans la base de donnée (Figure 3.2). L'apprentissage se termine lorsque l'algorithme atteint le critère d'arrêt défini pour notre système en 1000 itérations.

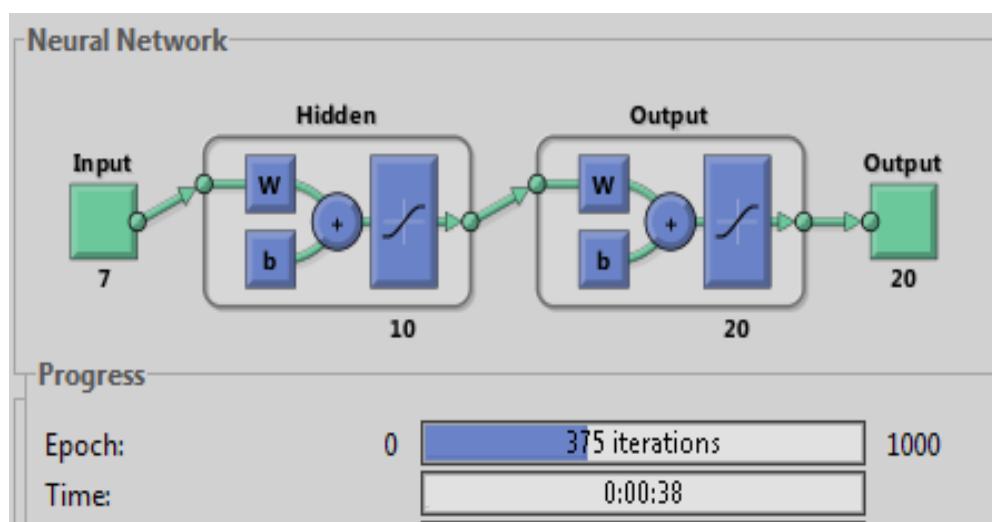


Figure 3.2 : Phase d'apprentissage du réseau de neurones

Une fois que le réseau présente des performances acceptables, il peut être utilisé pour une éventuelle phase de reconnaissance dans une deuxième étape. On a aussi, utilisé le descripteur les Moments de Legendre à l'ordre 3 qui nous a permis d'avoir 10 caractéristiques pour chaque visage de la base de données (ainsi la structure du réseau de neurones, pour ce cas, est 10 entrées indiquant les caractéristiques du moments de Legendre, 10 cellules de la couche cachée et 20 sorties correspondant aux nombre de visage dans la base de données) qui seront comparées avec ceux du visage détecté afin de prendre, une décision sur le résultat de la classification.

### 3.2.2.2. Les K-Plus proche voisin

L'instance  $x_q$  à classer, pour notre système le vecteur composé des caractéristiques du visage à détecter soit en utilisant les moments de Hu ou les moments de Legendre. Les  $x_1, x_2, \dots, x_k$  désignant les  $k$  instances à partir d'exemples d'apprentissage qui sont les plus proches de  $x_q$ , sont quatre qui correspondent au nombre de visages pour chaque classe de personne.

Différentes similarités ont été proposées, au cours des expériences, pour la classification par les K-plus proches voisins et seules les similarités par la distance Euclidienne et par la distance Manhattan ont été retenues pour la suite de notre projet.

### 3.2.3. Résultats et discussion

L'objectif de la première partie de ce chapitre est de présenter le système de reconnaissance adopté (figure 3.1) pour la reconnaissance d'un visage dans un document mixte, dans sa phase d'extraction et de classification. Nous présenterons les résultats trouvés dans chaque étape de cette phase du système.

Les documents présentés dans la figure 3.3 représentent les différents exemples traités, par notre système à l'aide d'une interface, développée sous MATLAB, qui propose un menu afin de choisir, pour chaque test, le descripteur (Moments de Hu ou Moments de Legendre) à utiliser au niveau d'extraction et le classificateur (K-plus proche voisins ou Réseaux de Neurones) à utiliser au niveau de la classification.

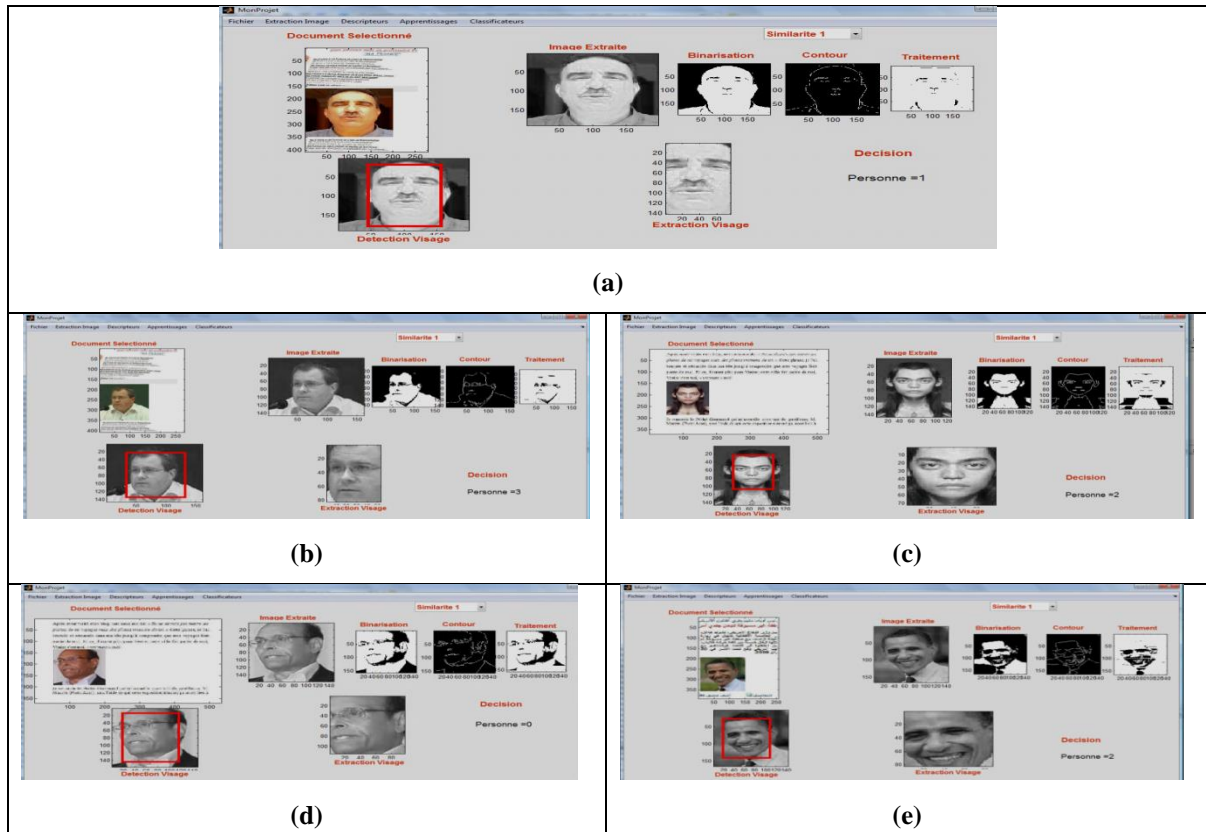


Figure 3.3 : Différents exemples de documents traités par la première approche proposée

Nous avons préparé une base de données pour les tests contenant soixante visages repartis sur cinq classes de personnes différentes à raison de douze visages variés, en fonction des expressions faciales de la personne et de la distance caractérisant l'image de la même personne au moment de sa prise, soit,  $(4 \times 3)$ , 4 visages (pour les différentes expressions) \* 3 (pour les différentes tailles de chaque visage). Ensuite, des résultats expérimentaux, basés sur des descripteurs particuliers, ont montré que ces descripteurs sont invariants, par conséquent la taille du visage (moyenne, petite ou grande) n'est pas un facteur à prendre en considération lors de la classification ce qui nous a ramené à réduire le nombre des visages dans la base de données, vu que les descripteurs à utiliser sont invariants, pour ne garder que vingt visages (figure 3.4), à raison de quatre visages pour chacune des différentes classes.



Figure 3.4 : Base de données utilisée pour les tests

Dans ce qui suit, nous présenterons les résultats expérimentaux obtenus par l'utilisation des moments de Hu et des moments de Legendre au niveau de la phase d'extraction d'attributs, ainsi que les réseaux de neurones et les K-Plus proches voisins au niveau de la phase de classification.

Le tableau 3.1 indique le Taux de Reconnaissance (TR) et le Taux d'Erreur (TE) calculés pour les différentes approches adoptées pour une base de données de vingt visages.

La première colonne indique le descripteur utilisé dans la phase d'extraction (le moments de Hu et de Legendre), la deuxième et la troisième colonne présentent le taux de reconnaissance et le taux d'erreur pour le cas des réseaux de neurones et les quatre dernières colonnes illustrent les résultats des taux de reconnaissance et d'erreur en adoptant le classificateur K-plus proche voisin pour les deux distances (Euclidienne ou Manhattan).

Descripteur	Réseau de Neurones		K-NN			
			Euclidien		Manhattan	
	T.R	T.E	T.R	T.E	T.R	T.E
Moments De Hu	70%	30%	65 %	35%	60 %	40%
Moments De Legendre	72%	28%	68%	32%	65%	35%

Tableau 3.1 : Résultats Expérimentaux

### 3.2.4. Illustration et interprétation des résultats

Nous avons effectué des tests sur des images que nous avons pris par un appareil-photo personnelle et d'autres sont téléchargées sur le Net. Les tests sont effectués sur des visages qui se présentent dans différentes couleurs et différentes orientations et échelles. Nous remarquons dans la :

Figure 3.3 (a), (b) et (c) la reconnaissance est positive car, pour chacun de ces trois cas, la décision est correcte du fait que chaque personne à identifier figure dans la base de données.

Figure 3.3 (d), là aussi la reconnaissance est positive car pour ce cas la décision est aussi correcte du fait que la personne à identifier ne figure pas dans la base de données.

Figure 3.3 (e), là contrairement au cas précédents, la reconnaissance est négative du fait que la personne à identifier ne figure pas dans la base de données alors que le système l'a identifié comme étant une personne de la base de données.

Les fausses reconnaissances de visages sont dues au fait que les caractéristiques extraites du visage d'une personne à détecter, qui ne fait pas partie des visages de la base de données, sont plus similaires à ceux du visage d'une personne qui fait partie des personnes dont les visages sont dans la base de données, à titre d'exemple le cas présenté dans la figure 3.3 (e).

## 3.3. Reconnaissance de visage :Deuxième approche proposée

### 3.3.1. Extraction des caractéristiques

L'étape d'extraction est l'étape clé dans le processus, car la performance de l'ensemble du système en dépend. Dans cette étape également connue sous le nom d'indexation ou de modélisation, est extraite des informations d'image du visage qui peuvent modéliser le visage d'une personne par un vecteur de mesure qui le caractérise (vecteur caractéristique ou signature).

Le choix judicieux de la méthode d'extraction est extrêmement important dans la mesure où la prochaine étape (classification) ne sera faite que sur la base des valeurs de

ces vecteurs. A cet effet nous avons pensé, dans l'étape d'extraction des caractéristiques du visage, à remplacer les moments invariants utilisés dans la première approche de cette étape, par une méthode géométrique (figure 3.5) basée sur des points caractérisant des régions importantes du visage comme c'est indiqué dans la figure 3.3 au niveau phase d'extraction.

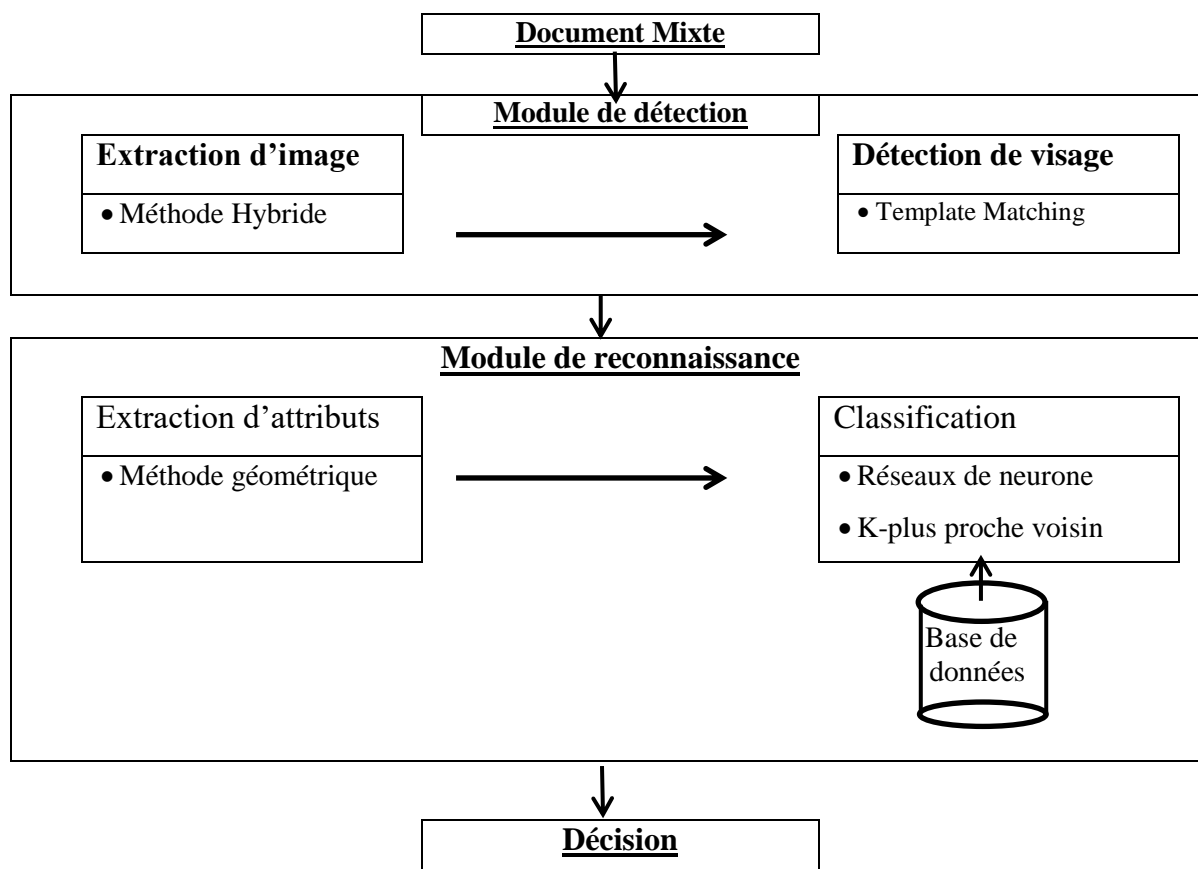


Figure 3.5 : Système élaboré en utilisant la deuxième approche proposée.

Les points caractéristiques utilisés sont choisis à base des résultats des recherches de la psychologie cognitive concernant le système visuel humain. Notamment le fait que la région des yeux, celle du nez et celle de la bouche sont importantes pour l'identification, les points qu'on utilise couvrent l'ensemble des trois régions auxquelles on a ajouté la région du front et celle des sourcils. Les caractéristiques, utilisées, par notre système, pour identifier un visage sont au nombre de 17 déduites à base 15 points (les 12 points du modèle WHO\_IS [Las 02], ayant un taux de reconnaissance de 86%, qui est à base du modèle de Kamel et al. [Kam 93] n'ayant utilisé que 9 points et qui avait un taux de reconnaissance de 66%. Et auxquelles on a rajouté 3 autres points

notamment les point centres des deux sourcils et un point milieu du franc comme c'est illustré dans la figure 3.6. Ainsi, on a jugé d'avance, que le taux de reconnaissance peut être amélioré avec des points de plus qui augmenteront légèrement le nombre de caractéristiques d'un visage [Bou 18].

A partir des 15 points déterminés par notre système, signalés auparavant, qu'on a déduit en utilisant la méthode 'Template Matching' comme c'est illustré dans la figure 3.4 présentée ci-dessous, on a constaté qu'on peut les classifier en deux types :

\*1<sup>er</sup> type : 11 positions invariantes. Il s'agit des positions suivantes :

- Les quatre positions P1, P2, P3, P4 et P12 déterminées par les extrémités de chacune des yeux plus celle du centre entre les yeux.
- Les positions P5, P6 et P7 déterminées par le nez.
- Les positions P8, P9 et P10 déterminées par la bouche.

\*2<sup>ème</sup> type : 4 autres positions, variantes, il s'agit des positions suivantes :

- Les positions P11, P13, P14 et P15 déterminées par la bouche et les sourcils à partir desquelles on a déterminé des distances invariantes ajoutées au vecteur caractéristique du visage utilisé pour la reconnaissance du visage.

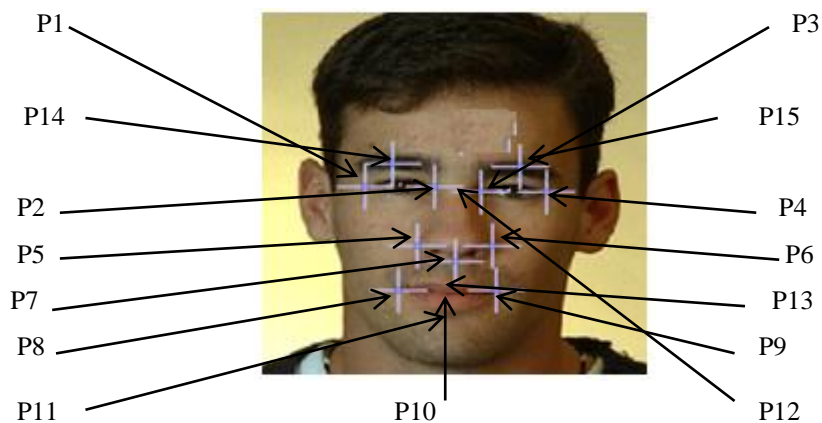


Figure 3.6 : Extraction des différentes positions exploitées par notre système

A partir des différentes positions, indiquées au paravent, on a déterminé donc 17 caractéristiques invariantes qu'on a considéré comme vecteur caractéristique du visage utilisé pour la reconnaissance du visage, il s'agit :

- Quartes caractéristiques : distances déterminées par les points des yeux et le point centre de la bouche (entre les positions P1, P2, P3 et P4 et la position P10);

- Quartres caractéristiques : distances déterminées par les points des yeux et le point bas du nez (entre les positions P1, P2, P3 et P4 et la position P7);
- Deux caractéristiques : distances déterminées par les points du nez et le point centre de la bouche (entre les positions P5 et P6 et la position P10);
- Une caractéristique : distances déterminée par les extrémités du nez (entre les positions P5 et P6);
- Deux caractéristique : distances déterminées par les points des yeux et le point milieu des deux yeux (entre les positions P1 et P2 la position P12);
- Une caractéristique : distances déterminée par l'axe centre de la tête (entre les positions P12 et P10);
- Une caractéristique : distances déterminée par les points de Cécile (entre les positions P13 et P14);
- Deux caractéristiques : distances déterminées par les points centre Franc et les points extrémités droite de l'œil gauche et extrémités gauche de l'œil droite (entre les points P15 et P2 puis entre P15 et P3).

### **3.3.2. Classification**

La classification est l'assignation d'une classe spécifique ou pas une classe de visage donnée: la classe ici est une personne avec des images de visage dans la base de données et la non-classe dans le cas où le visage n'appartient pas à aucune classe. Cette allocation nécessite l'introduction d'une mesure de similarité.

#### **Pour le cas des K-Plus proches voisins**

L'algorithme ne construit pas de modèle de classification mais est basé sur des valeurs trouvées dans le stockage ou la mémoire. Pour identifier la classe d'une entrée, l'algorithme sélectionne la classe à laquelle appartient la majorité des k plus proches voisins.

L'algorithme K-plus proches voisins est considéré comme l'un des algorithmes d'apprentissage automatique les plus simples. Cependant, il est coûteux en termes de calcul, en particulier lorsque la taille de l'ensemble d'apprentissage devient importante, ce qui rend la tâche de classification très lente [Bar 12].



Dans la deuxième version du système élaboré, nous avons gardé la valeur  $k=1$ . Ainsi, le plus proche visage est celui dont les caractéristiques sont les plus similaires à ceux de l'individu à identifier.

#### **Pour les réseaux de neurones**

De nombreuses applications s'intéressent à l'utilisation des réseaux de neurones, particulièrement pour la classification de données, la modélisation de processus complexes et le traitement non-linéaire des signaux. Les recherches ont engendré une panoplie d'architectures de réseaux dont chacune répond parfaitement à une application donnée.

Les réseaux de neurones ont été employés, dans le cas de la reconnaissance de visages dans tous les modules intervenant dans la chaîne de traitements. Ils sont utilisés aussi bien pour la détection de visages, pour l'extraction de signatures et pour la classification.

Les paramètres du réseau de neurones sont adaptés durant la phase d'apprentissage, afin d'obtenir en définitive une estimation simultanée des fonctions de séparation et de mélange. Le processus consiste à présenter l'une des images d'un individu en entrée et calculer la différence quadratique entre cette image et celle reconstruite en sortie qui correspond à l'erreur entre l'image initiale et l'image reconstruite par le réseau. En faisant ajuster les paramètres du réseau de façon itérative afin de minimiser l'erreur, on obtient en définitive un réseau qui peut extraire la signature la plus appropriée pour une classe donnée de visages.

Pour le cas de notre système, nous avons utilisé 17 neurones comme couche d'entrée représentant les caractéristiques déterminés par la méthode géométrique utilisée comme Approche d'extraction, 10 neurones pour la couche cachée et 40 neurones pour la couche de sorties correspondant aux nombre de classes de visages dans la base de donnée, comme c'est indiqué dans la table (Table 3.2) présentée ci-dessous.

Approche d'extraction	Couche d'entrée	Couche cachée	Couche de sortie
Méthode géométrique	17 neurones	10 neurones	40 neurones

Tableau 3.2. Caractéristiques du réseau de neurones utilisé par notre système.

### 3.3.3. Base de données

Notre système a été testé sur une grande partie de la base des données feret-color. Nous avons utilisé dans cette base de données 200 visages dont un extrait est illustré par la Figure 3.7. L'ensemble de ces visages est réparti sur 40 classes différentes de personnes à 5 faces variées chacune, en fonction des expressions faciales de la personne et de la distance caractérisant l'image de la même personne à sa portée.

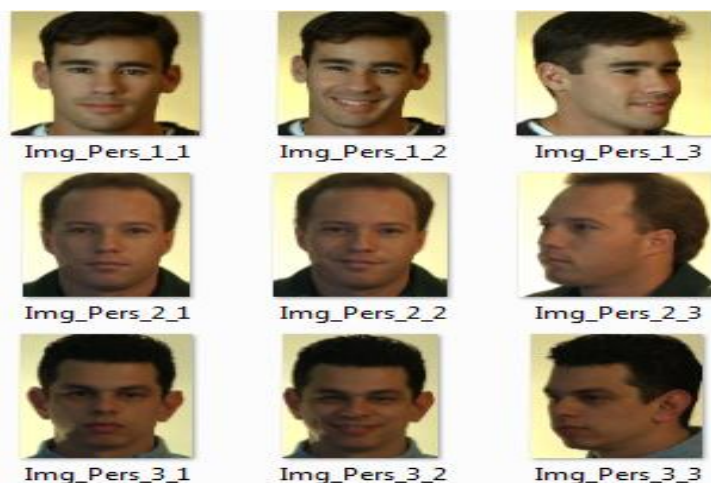


Figure 3.7 : Extrait de la base des données utilisée

Cette base de données est transformée au cours du traitement, en une base d'apprentissage dont le but est de réduire le temps de la phase de classification pour permettre à notre système de comparer, dans un bref délai, le visage en entrée avec tous les visages de la base de données utilisée.

## 3.4. Résultats et discussions

Au départ, nous avons effectué plusieurs tests, uniquement avec douze points de la méthode WHO\_IS. Le taux de réussite basés sur ces douze points était un peu moins proche du taux réalisé par la méthode WHO\_IS, à cause de la base des données utilisée par notre système qui est plus importante de point de vu nombre de visage en la comparant avec celle de la méthode WHO\_IS, ce qui nous a poussé à augmenter, d'avantage, le nombre de points pour atteindre quinze points à partir desquels on a déterminé dix-sept caractéristique pour former le vecteur caractéristique du visage ainsi les résultats sont largement améliorés comme l'indique le tableau 3.3 :

Descripteur	Réseau de neurones		K-plus proches voisins			
			Euclidien		Manhattan	
	R.R	E.R	R.R	E.R	R.R	E.R
Méthode géométrique	92%	8%	90%	10%	91%	9%

Tableau 3.3 : Résultats expérimentaux concernant la deuxième version du système

Après l'analyse des résultats réalisés par notre système, on a constaté qu'avec la méthode géométrique utilisée, pour chacun des classificateurs K-plus proches voisin et réseaux de neurones, que les résultats sont importants avec une légère faveur pour les réseaux de neurones.

## Conclusion

Au long de ce chapitre, on a détaillé les différentes étapes réalisées lors de l'implémentations de notre système qui sont traduites par la mise en place d'une interface qui donne une vue plus détaillée sur les différentes étapes et méthodes utilisées dans ce travail. Nous avons donc abordé plus précisément quelques aspects du problème. Tout d'abord, nous avons présenté la technique utilisée pour mesurer la similarité entre les images à partir des vecteurs descripteurs résultants de l'étape d'extraction des caractéristiques. Ensuite nous avons présenté une évaluation sur les résultats expérimentaux.

Enfin, à partir de ces résultats, nous allons établir, dans ce qui suit, une conclusion générale en citant quelques perspectives pour la suite de notre recherche.

## **Conclusion générale et perspectives**

L'objectif de cette thèse était de réaliser un système de reconnaissance d'un visage dans un document mixte. Ce processus devait s'inscrire dans un système complet de détection/reconnaissance traitant les informations provenant d'un document qui contient, aussi bien du texte que des images présentant ou non des visages dont le but est de ne détecter et identifier que les visages.

Nous nous sommes attachés à mettre au point une méthode de reconnaissance de visage rapide et fiable. Pour cela nous avons dans un premier temps choisi un descripteur d'images, avec pour objectif, qu'il soit suffisamment informatif mais peu coûteux en temps de calcul. Nous avons tout d'abord implémenté des techniques de détection de visage et des descripteurs classiquement utilisés dans ce domaine : les histogrammes, les moments de Hu et de Legendre. Les résultats n'étaient pas satisfaisants ; ce qui nous a poussé à développer d'autres techniques, aussi bien au niveau détection du visage qu'au niveau extraction du vecteur caractéristiques du visage détecté. Ainsi on a développé des algorithmes basés sur la couleur de la peau et la technique Template Matching pour la détection du visage et un descripteur adapté à notre cas, plus rapide en temps de calcul. C'est pourquoi nous avons exploré une nouvelle piste avec un descripteur géométrique.

Vient ensuite la tâche de classification qui utilise les informations fournies par le descripteur en amont. Là encore nous avons mis en œuvre les méthodes les plus populaires de cette dernière décennie : méthodes des k-plus proches voisins et les réseaux de neurones. Toutes ces méthodes ont été implémentées pour être testées et validées. De nombreuses combinaisons ont été expérimentées, en inter changeant les descripteurs et les classifieurs, ce qui a permis de mettre en évidence certains faits.

En ce qui concerne la détection de visage dans le document fourni en entré, nous avons testé deux approches : les histogrammes et une combinaison entre la technique basée sur la couleur de la peau et celle de la méthode Template Matching. C'est la technique impliquant les deux méthodes Hybride et Template Matching qui a les meilleures

performances pour cette application. Toutefois, l'autre descripteur obtient des résultats proches.

En ce qui concerne les descripteurs d'images, nous avons testé trois descripteurs: les moments de Hu et de Legendre et un descripteur géométrique basé sur des distances entre différents points invariants du visage. C'est le descripteur géométrique qui a les meilleures performances pour cette application. Toutefois, les deux autres descripteurs obtiennent des résultats proches.

Au terme de ce travail, le premier constat, trivial, est que ce type de challenge est très difficile. S'il nous paraît quasiment inné d'interpréter ce que nous voyons et d'agir en conséquence, la tâche est bien plus ardue à partir d'un système de vision.

La grande variabilité des visages (position, éclairage, expressions faciales...) est une grande difficulté pour entraîner un classifieur et trouver un modèle adéquat. Des améliorations pourraient certainement être apportées en utilisant des descripteurs plus performants ou en entraînant des classifieurs sur des bases d'apprentissage gigantesques. Mais l'utilisation de telles méthodes, plus riches en informations, nous est encore impossible pour des questions de temps de calcul. Il faut faire un compromis entre temps de calcul et performances.

---

## Références

- [Bap 15] Baptiste Chu, ‘Neutralisation des expressions faciales pour améliorer la reconnaissance du visage’, Ecole Centrale de Lyon, Ecole doctorale Informatique et Mathématiques, Thèse soutenu en Mars 2015.
- [Bar 12] F.Barigou, B. Atmani,Y.Bouziane, N.Barigou, ‘’ Accélération de la méthode des K plus proches voisins pour la catégorisation de textes’’, Laboratoire d’informatique d’Oran, Université d’Oran, BP 1524, El M’Naouer, Es Senia, 31 000 Oran, Algérie, 2012.
- [Bey 94] David J Beymer. Face recognition under varying pose. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR’94.*, 1994 IEEE Computer Society Conference on, pages 756–761. IEEE, 1994.
- [Ben 07] Ch. Bencheriet, A/H. Boualleg, H. Tebbikh, B. Guerzize and W. Belguidoum, ‘Détection de Visages par Méthode Hybride Couleur de Peau et Template Matching’, 4th International Conference, Sciences of Electronic, Technologies of Information and Telecommunications, Tunisia March 25-29, 2007.
- [Bou 15] L. Bouhou, R. El Ayachi, M. Fakir, M. Oukessou, ‘Recognition of a Face in a Mixed Document’, *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol. 15, No. 2, August 2015.
- [Bou 16] L. Bouhou, R. El Ayachi, M. Baslam, M. Oukessou, Face detection in a mixed-subject document, [*IJECE*] #12725, Vol. 6, No. 6, December 2016.
- [Bou 18] L. Bouhou, R. El Ayachi, M. Baslam, M. Oukessou, ‘Face Recognition in a Mixed Document based on the Geometric Method’, *International Journal of Advanced Science and Technology*, Vol.116 (2018), pp.109-120.
- [Cha 09] Ahmed Chaari, ‘Reconnaissance de visages par réseaux d’ondelettes de Gabor’, No d’ordre : 40181, Université Lille 1, décembre 2009.
- [Cha 10] Anis Chaari, ‘Nouvelle approche d’identification dans les bases de données biométriques basée sur une classification non supervisée’, Université d’Evry-Val d’Essonne, Français. <tel-00549395>, 2009, Submitted on Dec 2010.
- [Chi 03] C. Chiang, W. Tai, M. Yang, Y. Huang and C. Huang, ‘A novel method for detecting lips, eyes and faces in Real Time Imaging’, 9(4):277-287, 2003.

- [Cho 04] Chee-Way Chonga, P. Raveendranb and R. Mukundan, Translation and scale invariants of Legendre moments, *Pattern Recognition* 37, pp. 119–129, 2004.
- [Chu 15] B. Chu, ‘Neutralisation des expressions faciales pour améliorer la reconnaissance du visage’, Ecole Centrale de Lyon, 2015, <NNT : 2015ECDL0005>, HAL Id: <tel-01225809>.
- [Dai 98] D. Q. DAI, P. C. YUEN and G. C. FENG, ‘A multi-resolution decomposition method for human face recognition’, *Proceedings/actes Vision Interface* ', p. 301-307, Vancouver, British columbia, June 1998.
- [Ela 12] R. EL AYACHI, ‘Reconnaissance des Formes par une Classe d’Invariants et Applications à la reconnaissance des documents Tifinaghes’’, THESE, Volume 104, Pages 48-56, 2012.
- [Fic 12] C. Fiche, ‘Repousser les limites de l’identification faciale en contexte de vidéosurveillance’’, laboratoire Images Parole Signal et Automatique de Grenoble, Université de Grenoble, 2012.
- [Fre 00] C. Fredouille, J. Mariethoz, C. Jaboulet, J. Hennebert, J-F. Bonastre, C. Mokbel and F.Bimbot, ‘Behavior of a Bayesian Adaptation Method for Incremental Enrollment in Speaker Verification’, *International Conference on Acoustics, Speech, and Signal Processing*, pp. 1197-1200, Istanbul, Turquie, 5-9 Juin 2000.
- [Gar 99] C. Garcia and G. Tziritas, ‘Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis’, *IEEE Transactions on Multimedia*, p.264-277, September 1999.
- [Gue 08] S. GUERFI ABABSA, ‘Authentification d’individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D’, Université d’Evry-Val d’Essonne, 2008. HAL Id: tel-00623243, Submitted on Sep 2011.
- [Gom 02] G. Gomez, E. Morales, Automatic feature construction and a simple rule induction algorithm for skin detection, *Proceedings of Workshop on Machine Learning in Computer Vision*, pp. 31–38, 2002.
- [Haj 15] A. Hajraoui, ‘La reconnaissance de visages Humains’, Université Sultan Moulay Sliman, Centre d’Etudes Doctorales, Novembre 2015.
- [Heu 02] G. Heusch, ‘Détection automatique de visages dans une séquence vidéo’, SSC, Février 2002.

- [Hu 61] MING-KUEI HU "Visual Pattern Recognition by Moment Invariants", *Ire Transactions On Information Theory* vol. 49, pp. 830, January, 1961.
- [Jai 00] A. K. Jain, L. Hong and S. Pankanti, 'Biometrics : Promising Frontiers for Emerging Identification Market', *Communications of the ACM*, pp. 91-98, February 2000.
- [Jai 04] A. K. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition". *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4 – 20, Jan 2004.
- [Jon 99] M. J.Jones and J. M. Rehg, 'Statistical color models with application to skin detection', *Cambridge Research Series, CRL 98/11*, December 1998.
- [Kam 93] M. S. Kamel, H. C. SHEN, A. K. C. WONG and R. I. CAMPEANU, "System for the recognition of human faces", *IBM Systems Journal*, Vol. 32, No 2, p. 307-320, 1993.
- [Lam 16] F. Lamare, 'OCT en phase pour la reconnaissance biométrique parempreintes digitales et sa sécurisation', *Traitement du signal et de l'image*, Institut National des Télécommunications, 2016. France. <NNT: 2016TELE0003>.
- [Las 02] M. T. Laskri, Djallel Chefrour, "Who\_Is : système d'identification des visages humains", *Département d'Informatique , Université Badji Mokhtar Annaba, Algérie*, V.1 – 2002, p.39-69 ARIMA.
- [Ler 16] B. Leroy, A. Chouakria, I. Herlin, E. Diday 'Approche géométrique et classification pour la reconnaissance de visage', *Congrès de Reconnaissance des Formes et Intelligence Artificielle*, Rennes, France. HAL Id <inria-00532693> Apr 2016 .
- [Ley 10] Laetitia Leyrit, 'Reconnaissance d'objets en vision artificielle : application à la reconnaissance de piétons', *Université Blaise Pascal - Clermont-Ferrand II*, N° d'ordre : D.U 2071, EDSPIC : 498, 2010, Submitted on Sep 2011.
- [Li Tian 01] Ying liTian, TakeoKanadeet Jeffrey F.Cohn.RecognizingAction Units for Facial Expression Analysis.*IEEE Transactions on PatternAnalysis and Machine Intelligence*, 23(2) :97–115, 2001.
- [Liu 01] S. Liu and M. Silvean, 'A pratical Guide to Biometric Security Technology', *IEEE Computer Society, IT Pro-Security*, Janvier-Février 2001.
- [Mer 07] H. Mercier, 'Modélisation et suivi des déformations faciales', *Université Paul Sabatier - Toulouse III, Français*. <tel-00185084> 2007.



- [Ouj 12] M. Oujoura, B. Minaoui and M. Fakir, 'Multilayer Neural Networks and Nearest Neighbor Classifier Performances for Image Annotation', (IJACSA), International Journal of Advanced Computer Science and Applications, Vol.3, No.11, 2012.
- [Per 02] F. Perronnin and J. Dugelay, 'Introduction à la Biométrie Authentification des Individus par Traitement Audio-Vidéo Traitement du Signal', Volume 19 –n° 4, 2002.
- [Sin 94] P. Sinha, 'Object Recognition via Image Invariants: A Case Study, Investigative Ophthalmology and Visual Science', vol. 35, no. 4, pp. 1735-1740, 1994.
- [Tea 80] M.R. Teague, Image analysis via the general theory of moments, J. Opt.Soc. Amer. 70 , pp.920–930, 1980.
- [Tou 05] S. TOUHAMI, Classification non supervisée de pixels d'images couleur par analyse d'histogrammes tridimensionnels, Laboratoire d'Ingénierie des Systèmes Automatisés, Université d'Angers, septembre 2005.
- [Tur 91] M. A. TURK and A. P. PENTLAND, 'Face recognition using eigenfaces', IEEE Comput. Sco.Press, p. 586-591, June 1991.
- [Pee 03] P. Peer, J. Kovac and F. Solina, 'Human skin colour clustering for face detection', In submitted to Eurocon 2003 – International Conference on Computer as a Tool, 2003.
- [Vio 01] P. Viola, M. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, Conference On Computer Vision And Pattern Recognition 2001].
- [Yan 02] M. Yang, D. Kriegman and N. Ahuja, 'Detecting faces in images : A survey. IEEE Transactions on pattern analysis and machine intelligence, 24(1):1746 -1762, 2002.
- [Zhe 09] W. Zheng and S. Bhandarkar, 'Face detection and tracking using a boosted adaptive particle filter', Journal of Visual Communication and Image Representation, 20(1):9 - 27, 2009.