الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد د حلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique

Mémoire de Master

Mention Électronique
Spécialité Electronique des systèmes embarqués ESE

présenté par

Hadj Attou Abdelkader

&

BOUIZZOUL ABD ELDJALAL

Simulation d'un système d'identification de personnes par le visage et la voix

Proposé par : GUESSOUM Abderrezak&H.bougherira

Année Universitaire 2017-2018

Remerciements

Au début, nous remercions Dieu Tout-Puissant de nous avoir donné patience, santé et volonté.

Nous voudrions présenter nos remerciements à notre encadreur « P. GUESSOUM ». Nous voudrions également lui témoigner notre gratitude pour sa patience et son soutien qui nous a été précieux afin de mener notre mémoire à bon port.

Aussi, on tient à remercier Mme H.bougherira pour son orientation et ces efforts afin qu'on puisse mener à bien ce mémoire.

Nous tenons également à remercier tous les enseignants du département électronique de Blida qui ont partagé leur savoir avec nous.

Nous remercie également tous les membres du jury d'avoir accepté d'assister à la présentation de cemémoire.

Sans oublier de remercier nos parents pour tous les sacrifices qu'il a faits pour nous.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.

ملخص:

تقدم هذه المذكرة رؤية جديدة للأمن حيث يمكن التعرف على شخص من وجهه وصوته. التعرف على الوجه هو أسلوب مهم وصعب للغاية للتعرف التلقائي بالناس. في هذا العمل يتم استخدام طريقة فيولاوجونز كخطوة للكشف عن الوجه،يتم استخدام استخراج السمات المحلية من الرسم البياني من التدرج الموجه (HOG) للعب دور رئيسي في وظيفة تحديد الهوية.لقد قمنا بتكييف تقنية تكميم المتجهات وتقنية معامل ميل للتردد الطفي (MFCC)للتعرف التلقائي على المتحدث.يتم تنفيذ الدمج من أجل الجمع بين القرارات التي يقدمها النظامان مع نهج اندماج عالي المستوى يستند التصويت إلى الجمع المنطقي.

كلمات المفاتيح:HOG ، تكميل المتجهات، MFCC .

Résumé : Ce mémoire présente une vision nouvelle de la sécurité où il serait possible d'identifier un individu par l'acquisition et connaissance de son visage et sa voix. Dans ce travail, La méthode de Viola et Jones est utilisée comme une étape de détection de visages, l'extraction des caractéristiques locales de HOG est utilisée pour jouer un rôle clé dans la fonction d'identification. Nous avons adapté la technique de quantification vectorielle et la technique MFCC pour la reconnaissance automatique du locuteur. La fusion est mise en œuvre pour combiner les décisions prises par les deux systèmes avec une approche fusion de haut niveau basée sur le vote ET.

Mots clés :HOG;quantification vectorielle ; MFCC.

Abstract: This thesis presents a new vision of securitywhere a person can beidentified by his face and hisvoice. Facial recognition is an important and verydifficult technique for automatic recognition of people. In thiswork, The Viola and Jones methodisused as a face detectionstep, the extraction of the local features of HOG isused to playa key role in the identification function. We have adaptedvectorquantization technique and MFCC technique for automatic speaker recognition. The merge isimplemented to combine decisions made by the twosystemswith a high-level fusion approachbased on the AND voting.

Keywords : HOG, VectorQuantization, MFCC.

Listes des acronymes et abréviations

ADL Analyse Discriminante Linéaire

HOG Histogramme de gradient orienté

VAL Vérification Automatique de Locuteur

RAL Reconnaissance Automatique du Locuteur

IAL Identification Automatique de Locuteur

FFT Transformation de Fourier rapide

DCT Transformée en cosinus discrète

MFCC Mel Frequency Cepstral Coefficient

AR autorégressive

PLP Prévision linéaire perceptuelle

LPC Codage prédictif linéaire

DTW Déformation temporelle dynamique

QV Quantification vectorielle

HMM Modélisation de Markov cachée

GMM Modèle de mélange gaussien

SVM Soutenir les machines vectorielles

LBG Linde, Buzo, Gray

DE distance euclidienne

PDFProbability Density Function

BD base de données

FA Fuse Acceptation

FR Faux rejet

ACP Analyse des Composantes Principales (EigenFace)

ADL Analyse Discriminante Linéaire (FisherFace)

Table des matières

Liste des figures

Liste des tableaux

Introduction générale

Introduction générale

La sécurité des systèmes d'information est devenue un domaine de recherche d'une très grande importance. La conception d'un système d'identification fiable, efficace et robuste est une tâche prioritaire. L'identification de l'individu est essentielle pour assurer la sécurité des systèmes et des organisations. Elle correspond à la recherche de l'identité de la personne qui se présente dans une base de données et peut servir à autoriser l'utilisation des services.

Chaque être humain peut, dès son plus jeune âge, reconnaître les voix et les visages des personnes qui lui sont familières. En fait, comment peut-on de manière automatique par ordinateur reconnaître un individu par la seule prise en compte de l'image de son visage ou à partir d'un échantillon de sa voix ? La réponse à cette seule question sera le fil directeur de notre travail.

Le visage et la voix peut être considéré comme des données biométriques. Une donnée biométrique est une donnée qui permet l'identification d'une personne sur la base de ce qu'il est (caractéristiques physiologiques ou comportementales). Les indices biométriques physiologiques sont des traits biologiques/chimiques innés, alors que les indices biométriques comportementaux sont associés à des comportements appris ou acquis.

Les données biométriques sont devenues des données incontournables pour le problème de l'identification sécurisée et de la vérification de personnes. Les méthodes d'identification ou de vérification d'identité basées sur une donnée biométrique offrent certains avantages par rapport aux méthodes basées sur un mot de passe ou un code PIN.

Notre travail s'intéresse à la fusion de la reconnaissance faciale et de la reconnaissance automatique du locuteur (La fusion visage voix).Elle consiste à réaliser un Système de Reconnaissance faciale basé sur la méthode de Viola et Jones et l'extraction des caractéristiques locales de HOG, un Système de Reconnaissance Automatique de locuteur à base de MFCC et la quantification vectorielle (QV).Ce système s'appuie sur une combinaison série de classificateurs de différente types, et utilise comme

méthode de fusion l'approche par combinaison basée sur le vote AND (si tous les systèmes ont décidé 1 alors OUI) pour fusionner les décisions.

Ce document est organisé comme suit : dans le premier chapitre, nous présentons l'état de l'art d'un système de reconnaissance faciale, les principales méthodes utilisées pour la détection et l'identification de visages, les méthodes choisies.

Dans le chapitre 2 nous présentons l'état de l'art d'un système de reconnaissance automatique du locuteur, les principales méthodes utilisées et les méthodes choisies.

Dans le chapitre 3 nous présentons la fusion et ses différents niveaux, les principales méthodes utilisées, la méthode choisie.

Dans le chapitre 4 nous présenterons le logiciel utilisé pendant la conception et la réalisation de notre solution et afin nous terminerons par les résultats obtenus et commentaires observés.

Chapitre1 :Reconnaissance faciale

1.1 Introduction

Depuis des années La reconnaissance faciale été un sujet de recherche dans le domaine de la vision par ordinateur. Maintenant la reconnaissance faciale est devenue une réalité, les systèmes de reconnaissance faciale sont les plus communes et populaires.

La reconnaissance faciale est une technique d'identification biométrique fondée sur un traitement automatique d'images numériques d'un individu et permettant de l'identifier à partir des caractéristiques de son visage. Inspirée par le fonctionnement de l'œil humaine. Le principe est simple : un capteur « saisit » un visage, le transforme en données numériques pour ensuite le comparer à une base de données, ces deux dernières opérations étant réalisées par un algorithme.

La reconnaissance faciale permet d'adapter la vérification biométrique à toutes les situations, C'est une technologie très efficace qui est utilisée dans de nombreuses applications liées à la sécurité. Elle est par exemple un outil très fiable pour aider les forces de police à identifier des criminels, ou bien pour permettre aux services de douanes de vérifier l'identité des voyageurs. Actuellement, avec la numérisation des échanges, l'usage de cette technologie est en train de s'étendre au monde des entreprises. Utilisée dans des applications commerciales, la reconnaissance faciale permet par exemple de sécuriser des transactions en ligne. La reconnaissance faciale est sans contact et son utilisation ne nécessite aucun outil spécifique, ce qui en fait la solution idéale pour l'identification de personnes dans une foule ou dans des espaces publics [1].

La reconnaissance faciale est une aptitude qui relie l'apparence d'une personne à son identité. Lors d'une rencontre, cette compétence permet de se rappeler des échanges précédents et ainsi de construire une relation à long terme où les individus finissent par se connaître et savoir anticiper leurs comportements et besoins respectifs [2].
La reconnaissance faciale est donc une technologie qui a atteint une certaine maturité au point d'en faire un outil non seulement de la vie quotidienne mais également une solution parmi d'autres pour améliorer la sécurité. L'enjeu est celui des normes techniques qui garantissent un niveau de qualité et d'exploitation large. C'est aussi un enjeu économique pour le développement des produits et leur exploitation.

1.2 Présentation du système de reconnaissance faciale

Les systèmes de reconnaissance faciale sont des systèmes automatisés capables d'identifier des individus en fonction des caractéristiques de leur visage telles que l'écartement des yeux, des arêtes du nez, des commissures des lèvres, des oreilles, menton, etc. Ces caractéristiques sont analysées puis comparées à une base de données existante afin d'identifier une personne ou de vérifier son identité.
Dans un système de reconnaissance faciale, une image suit depuis son entrée un processus bien précis pour arriver à déterminer l'identité du porteur de visage. Ce processus comporte plusieurs étapes qui peuvent être illustrées par le schéma suivant:

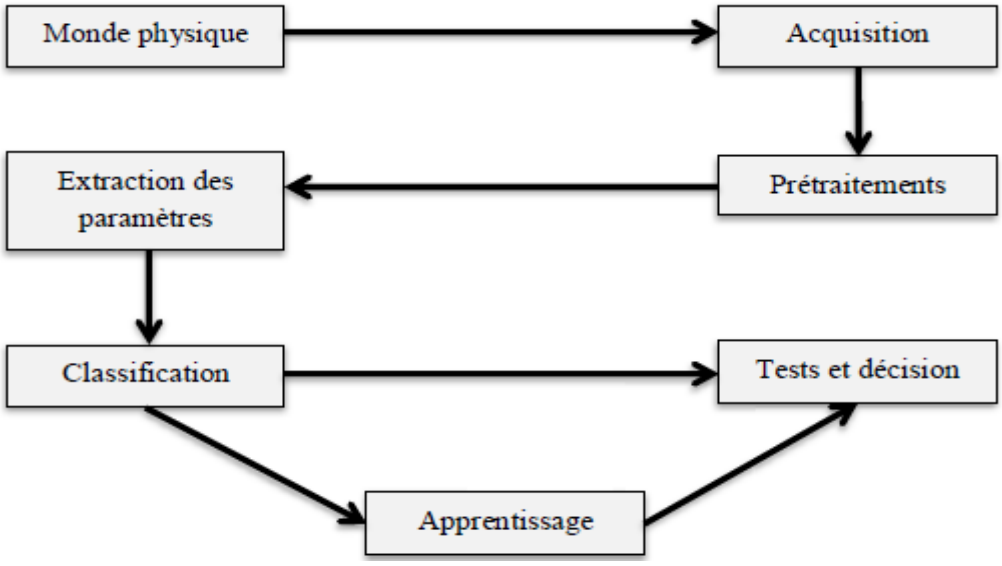

Figure 1. 1. *Processus d'un système de reconnaissance faciale [4].*

Donc pour être identifié, l'image d'une personne dans un système de reconnaissance faciale suit le processus suivant :

1.2.1 Le monde physique : (l'extérieur)

C'est le monde réel en dehors du système avant l'acquisition de l'image. Dans cette étape, on tient compte généralement de trois paramètres essentiels : L'éclairage, la variation de posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents, et par conséquence une fausse identification.

1.2.2 L'acquisition de l'image :

C'est l'opération qui permet d'extraire du monde réel une représentation matricielle c'est-à-dire une image, cette opération peut être statique (Appareil photo, Scanner, etc.) ou dynamique (Caméra, Web Cam).

1.2.3 Prétraitements :

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

1.2.4 L'extraction des paramètres :

En plus de la classification, l'étape de l'extraction des paramètres représente le cœur du système de reconnaissance, elle consiste à effectuer le traitement de l'image dans un autre espace de travail plus simple et qui assure une meilleure exploitation de données, et donc permettre l'utilisation, seulement, des informations utiles, discriminantes et non redondantes.

1.2.5 La classification : (modélisation)

Cette étape consiste à modéliser les paramètres extraits d'un visage ou d'un ensemble de visages d'un individu en se basant sur leurs caractéristiques communes. Un modèle est un ensemble d'informations utiles, discriminantes et non redondantes qui caractérise un ou plusieurs individus ayant des similarités.

1.2.6 L'apprentissage :

C'est l'étape où on fait apprendre les individus au système, elle consiste à mémoriser les paramètres, après extraction et classification, dans une base de données bien ordonnées pour faciliter la phase de reconnaissance et la prise d'une décision, elle est en quelque sorte la mémoire du système.

1.2.7 La décision :

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

1.3 Problématique

Le problème de la reconnaissance de visages peut être formulé comme suit : étant données une ou plusieurs images d'un visage, la tâche est de trouver ou de vérifier l'identité d'une personne par comparaison de son visage à l'ensemble des images de visage stockées dans une base de données.
Nous nous limitons dans ce travail à une reconnaissance à partir d'une image 2D de visage en environnements non contraints. De tels systèmes doivent pouvoir s'affranchir des problèmes suivants :

1.3.1 Influence des variations de la pose :

Les changements d'orientation et les changements de l'angle d'inclination du visage engendrent de nombreuses modifications d'apparence dans les images collectées.

Les rotations en profondeur engendrent l'occultation de certaines parties du visage comme pour les vues de trois-quarts. D'autre part, elles amènent des différences de profondeur qui sont projetées sur le plan 2D de l'image, provoquant des déformations qui font varier la forme globale du visage. Ces déformations qui correspondent à l'étirement de certaines parties du visage et la compression d'autres régions font varier aussi les distances entre les caractéristiques faciales.

1.3.2 Influence des changements d'éclairage :

L'intensité et la direction d'éclairage lors de la prise de vue influent énormément sur l'apparence du visage dans l'image. En effet, dans la plupart des applications courante, des changements dans les conditions d'éclairage sont inévitables, notamment lorsque les vues sont collectées à des heures différentes, en intérieur ou en extérieur.
Etant donnée la forme spécifique d'un visage humain, ces variations d'éclairage peuvent y faire apparaître des ombres accentuant ou masquant certaines caractéristiques faciales.

1.3.3 Influence des expressions faciales :

Les visages sont des éléments non rigides. Les expressions faciales véhiculant des émotions, combinées avec les déformations induites par la parole, peuvent produire des changements d'apparence importants, et le nombre de configurations possibles est trop important pour que celles-ci soient décrites in extenso de façon réaliste.

1.3.4 Influence des occultations :

Un visage peut être partiellement masqué par des objets ou par le port d'accessoires tels que lunettes, un chapeau, une écharpe. Les occultations peuvent être intentionnelles ou non. Dans le contexte de la vidéosurveillance, il peut s'agir d'une volonté délibérée d'empêcher la reconnaissance. Il est clair que la reconnaissance sera d'autant plus difficile que peu d'éléments discriminants seront simultanément visibles.

1.4 Domaines d'application

On retrouve La reconnaissance faciale dans plusieurs domaines :

a-Sécurité de l'information *:* sécurité des bases de données, cryptage de fichiers, Sécurité de l'intranet, accès à internet, s'enregistrer sur une installation personnelle.

b- **Droit d'accès et la surveillance** : Vidéo surveillance avancée, contrôle d'accès, analyse d'événements, poursuite des suspects et investigation.

c-Sécurité : dans les stades, les aéroports et les centres commerciaux dans plusieurs pays pour interdire l'accès de certains individus fichés.

1.5 Avantage et inconvénient de reconnaissance faciale

Avantages	Inconvénients
Technologie bien acceptée par le public	Technologie sensible à l'environnement (éclairage, expression du visage).
En position fixe et éclairée, les taux de reconnaissance sont effectivement très élevés	Technologie sensible au changement (barbe, moustache, chirurgie, perçage...).
Technique peu coûteuse	Les vrais jumeaux ne sont pas identifiés.

Tableau 1. 1.Avantage et inconvénient de Reconnaissance faciale

1.6 Méthodes de détection des visages :

La détection de visage dans l'image est un traitement indispensable et crucial avant la phase de reconnaissance. En effet, le processus de reconnaissance de visages ne pourra jamais devenir intégralement automatique s'il n'a pas été précédé par une étape de détection efficace.

Le traitement consiste à rechercher dans une image la position des visages et de les extraire sous la forme d'un ensemble d'imagettes dans le but de faciliter leur traitement ultérieur. Un visage est considéré correctement détecté si la taille

d’imagette extraite ne dépasse pas 20% de la taille réelle de la région faciale, et qu’elle contient essentiellement les yeux, le nez et la bouche.

L'intérêt de la localisation faciale va au-delà de l'application de ce présent mémoire. Son utilité se manifeste dans des domaines variés allant de la vidéosurveillance au jeu interactif. Les premières difficultés rencontrées par les méthodes s'attelant à détecter les visages sont les variations de pose d'expression, de rotation du visage, d'âge et d'illumination [2]. Pour le reste la difficulté est d'autant plus grande que la plupart des applications ayant recours à cette technologie requièrent une exécution en temps réel, limitant les marges de manœuvre de l'algorithme.

Les méthodes sont divisées en quatre catégories. Ces catégories peuvent se chevaucher si un algorithme peut appartenir à deux ou plusieurs catégories. Cette classification peut être faîte comme suit :

1.6.1 Méthodes basées sur les connaissances:

Ces méthodes se basent sur la connaissance des différents éléments qui constituent un visage et des relations qui existent entre eux. Ainsi, les positions relatives de différents éléments clés tels que la bouche, le nez et les yeux sont mesurées pour servir ensuite à la classification ’visage’ ’non visage’ chez Chiang et al [5]. Le problème dans ce type de méthode est qu’il est difficile de bien définir de manière unique un visage. Si la définition est trop détaillée, certains visages seront ratés tandis que si la description est trop générale, le taux de faux positifs montera en flèche.

1.6.2 Approches invariantes des caractéristiques:

Ces approches utilisent les éléments invariants aux variations d’illumination, d’orientation ou d’expression tels que la texture ou la signature de couleur de la peau pour la détection.

1.6.3 Méthodes de correspondance de modèle:

Les templates peuvent être définis soit " manuellement", soit paramétrés à l’aide de fonctions. L’idée est de calculer la corrélation entre l’image candidate et le template.

Ces méthodes rencontrent encore quelques problèmes de robustesse liés aux variations de lumière, d'échelle, etc. les invariants aux changements de luminosité permettant de caractériser les différentes parties du visage.

1.6.4 Méthodes basées sur l'apparence :

Ces approches appliquent généralement des techniques d'apprentissage automatique. Ainsi, les modèles sont appris à partir d'un ensemble d'images représentatives de la variabilité de l'aspect du visage. Ces modèles sont alors employés pour la détection. Ces méthodes présentent l'avantage de s'exécuter très rapidement mais demandent un long temps d'entraînement. Les méthodes appartenant à cette catégorie ont montré de bons résultats par rapport aux trois autres types de méthodes [6].On peut citer parmi celles-ci, la méthode basée sur les réseaux de neurones de Rowley et al, la méthode de Schneiderman et Kanade basée sur un classifieur de Bayes naïf ainsi que le fameux algorithme de Viola et Jones fonctionnant en temps réel.

1.7 Méthodes choisies

La méthode choisie est basée sur l'apparence ("Appearance-basedmethods") qui est le fameux algorithme de Viola et Jones fonctionnant en temps réel qui sera détaillé plus loin. Il existe d'autres méthodes mais celle de Viola et Jones est la plus performante à l'heure actuelle.

1.8 La méthode Viola et Jones

La méthode de Viola et Jones 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.

En tant que procédé d'apprentissage supervisé, la méthode de Viola et Jones nécessite de 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. La renommée de cette approche est faite sur trois concepts :

1.8.1 L'image intégrale

L'algorithme se base sur les caractéristiques de Haar (Haarfeatures) pour localiser les visages présents sur une image d'entrée. Dans le but d'extraire rapidement ces caractéristiques, l'image est représentée sous forme intégrale. En effet, sous cette forme, l'extraction d'une caractéristique à n'importe quel endroit et à n'importe quelle échelle est effectuée en un temps constant tandis que le temps de conversion vers la représentation intégrale ne remet pas en cause ce gain de temps offert par l'utilisation de la représentation en image intégrale. La définition des caractéristiques de Haar et la manière dont la représentation intégrale accélère considérablement leur extraction sont présentés ci-après pour une image en niveaux de gris [2].
Plus formellement, l'image intégrale iiau point $(x;\ y)$est définie à partir de l'image $i$par :

$$ii(x,y) = \sum_{x' \leq x,\ y' \leq y} i(x',y') \qquad \textbf{(1.1)}$$

Présenté comme tel, le calcul d'une caractéristique de Haar demande à chaque fois l'accès aux valeurs de tous les pixels contenus dans les zones rectangulaires considérées. Cela devient vite contraignant temporellement dès que les caractéristiques de Haar sont définies par des zones rectangulaires de grandes dimensions. L'image intégrale permet de surmonter ce problème en rendant constant le temps de calcul d'une caractéristique de Haar à n'importe quelle échelle. Dans toute image, une zone rectangulaire peut être délimitée et la somme des valeurs de ses pixels calculée. Une caractéristique de Haar est une simple combinaison linéaire de sommes ainsi obtenues.

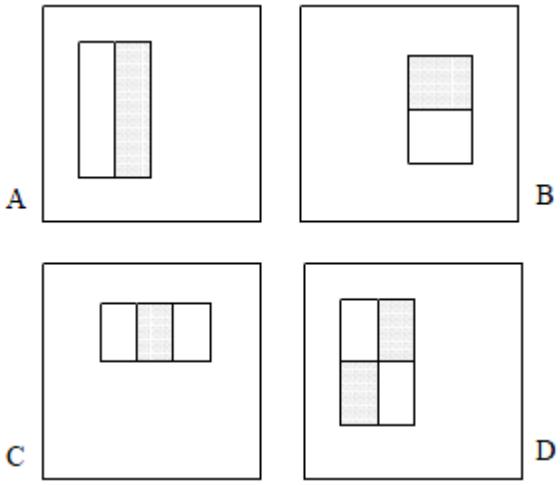

Figure 1. 2.Exemple de 4 caractéristiques de Haar.

Le calcul de la somme des valeurs des pixels appartenant à une zone rectangulaire s'effectue donc en accédant seulement à quatre pixel de l'image intégrale : soit un rectangle ABCD dont les sommets sont nommés dans le sens des aiguilles d'une montre en commençant par le sommet supérieur gauche et soit x la valeur sous la représentation intégrale d'un sommet X du rectangle($X \varepsilon\{A; B; C; D\}$). La somme des valeurs des pixels appartement à ABCD est, quelle que soit sa taille, donnée par c -b -d + a.

Une caractéristique de Haar étant une combinaison linéaire de tels rectangles ABCD, son calcul se fait alors en un temps indépendant de sa taille.

1.8.2 Algorithme d'apprentissage basé sur Adaboost

Pour localiser les visages sur l'image d'entrée, cette dernière est scannée par une fenêtre de dimension déterminée. La fenêtre parcourt l'image et son contenu est analysé pour savoir s'il s'agit d'un visage ou non. Comme dit plus haut, les caractéristiques de Haar sont extraites pour effectuer la classification et de ce fait la représentation intégrale de l'image accélère l'analyse. Mais, pour une fenêtre de 24x24 pixels il y a 45 396 caractéristiques de Haar, les traiter toutes prendrait beaucoup trop de temps pour une application en temps réel. Pour surmonter ce problème, une variante de la méthode de boostingAdaboost est utilisée.

Adaboost est une méthode d'apprentissage permettant de "booster" les performances d'un classifieur quelconque nommé "classifieur faible". L'idée est de faire passer les candidats à classifier à travers plusieurs classifieurs faibles, chacun étant entrainé en

portant plus d'attention sur les candidats mal classifies par le classifieur précédent. Pour arriver à ce résultat des poids sont associés aux échantillons du set d'entrainement ((*xi*;*yi*) *i*= 1; …; m), tout d'abord de manière équilibrée :

$$w_i^0 = \frac{1}{m}$$

Pour$i = 1; \ldots m$. Le 0 en exposant indique qu'il s'agit des poids initiaux. Ensuite le premier classifieur faible est entrainé comme suit :

$$h^0 = argmin_{h_j \in H} \epsilon_j$$

Avec l'erreur $\epsilon_j = \sum_{i=0}^{m} \omega_i^0 \delta(y_i - h_j(x_i))$et H l'ensemble des classifieurs faibles. Puis une nouvelle génération de poids w_i^1sont créés tels qu'ils accordent plus d'importance aux échantillons mal classifies par h_0. Ensuite un nouveau classifieur $h_1$est entrainé, puis de nouveaux poids w_i^2sont générés et ainsi de suite. Enfin, après T itérations, le classifieur fort H(x) est obtenu :

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x) - \theta\right) \quad \textbf{(1.2)}$$

Avec $\alpha = \frac{1}{2} ln \frac{1-\epsilon t}{\epsilon t}$et θ un seuil à déterminer. La valeur à attribuer à ce dernier sera discutée plus loin. Chaque classifieur fort est donc constitué d'un nombre T de classifieursfaibles.

Adaboost sert donc à booster un classifieur déjà existant et à priori chaque classifieur faible possède le même espace d'entrée. Dans la variante d'Adaboost de Viola & Jones, les classifieurs faibles $hj \in H$ ont pour entrée une caractéristique de Haar différente. Adaboost s'apparente alors à une sélection de caractéristiques (featureselection). Cette variante d'Adaboost est utilisée lors de l'apprentissage pour sélectionner les caractéristiques de Haar les plus à même de détecter un visage et permet ainsi de surmonter le problème du nombre élevé de caractéristiques de Haar existant pour une fenêtre de recherche.

1.8.3 Cascade de classifieurs

La méthode de Viola et Jones est basée sur une approche par recherche exhaustive sur l'ensemble de l'image, qui teste la présence de l'objet dans une fenêtre à toutes les positions et à plusieurs échelles. Cette approche est cependant extrêmement coûteuse en calcul. L'une des idées-clés de la méthode pour réduire ce coût réside dans l'organisation de l'algorithme de détection en une cascade de classifieurs. Appliqués séquentiellement, ces classifieurs prennent une décision d'acceptation ; la fenêtre contient l'objet et l'exemple est alors passé au classifieur suivant, ou de rejet ; la fenêtre ne contient pas l'objet et dans ce cas l'exemple est définitivement écarté. L'idée est que l'immense majorité des fenêtres testées étant négatives (c.-à-d. ne contiennent pas l'objet), il est avantageux de pouvoir les rejeter avec le moins possible de calculs. Ici, les classifieurs les plus simples, donc les plus rapides, sont situés au début de la cascade, et rejettent très rapidement la grande majorité des exemples négatifs. Cette structure en cascade peut également s'interpréter comme un arbre de décision dégénéré, puisque chaque nœud ne comporte qu'une seule branche [7].

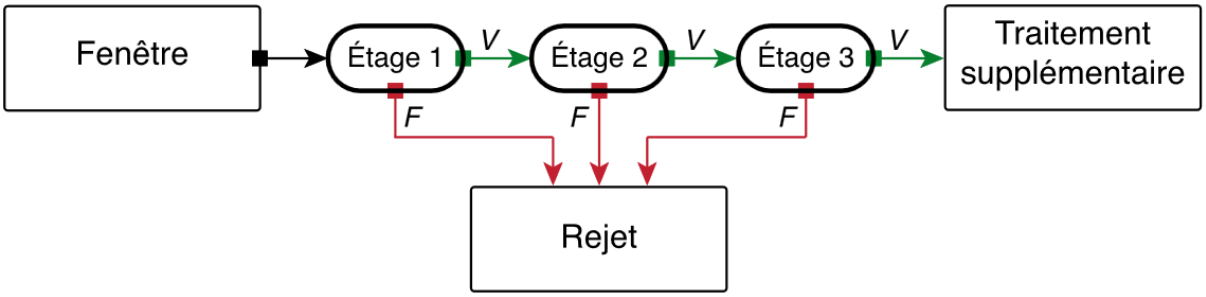

Figure 1. 3.Illustration de l'architecture de la cascade.

1.9 Méthodes de reconnaissance des visages

Les systèmes de reconnaissance de visages sont très souvent classés à partir des conclusions d'études psychologiques sur la façon dont les hommes utilisent les caractéristiques faciales pour reconnaitre les autres. De ce point de vue, on distingue les trois catégories : les méthodes globales, les méthodes locales et les méthodes hybrides.

1.9.1 Méthodes globales

Ces méthodes dites globales ou encore holistiques sont des méthodes qui utilisent toute la région du visage comme information d'entrée. Son principal inconvénient peut résider dans la taille des données à stocker lors de la phase d'apprentissage. Cependant avec l'actuel capacité de stockage de nos ordinateurs, nous pouvons relativement palier à ce problème.
Parmi les méthodes globales, nous pouvons citer :
_ L'ACP Analyse des Composantes Principales (encore appelée EigenFace).
_ L'ADL Analyse Discriminante Linéaire (encore appelée FisherFace) : les méthodes basées sur l'Analyse Discriminante Linéaire (ADL) déterminent les directions de projection les plus discriminantes dans l'eigenspace. Pour cela, elles maximisent les variations inter personne par rapport aux variations intra-personne. Ainsi, la méthode basée ADL se réduit alors à la méthode eigenface.
_ L'approche Probabiliste transforme le problème d'identification de visage en un problème de classification en deux classes. Elle évalue la probabilité de la différence entre une image de test et une image prototype appartenant aux classes intra-personne et inter-personne. Notons que la distribution intra-personne ne peut pas être évaluée dans le cas d'un exemple par personne, et la méthode se ramène aussi à la méthode eigenface.

1.9.2 Méthodes locales

Cette catégorie se divise en deux sous-catégories basées respectivement sur des caractéristiques et des apparences locales.

a-Méthodes basées sur des caractéristiques locales

Ici des propriétés géométriques sont extraites à partir de la localisation de points clés sur le visage. Deux faiblesses en découlent directement : d'une part la localisation de tels points n'est pas toujours une mince affaire lorsque des occlusions ou des variations de position ou d'expression surviennent et d'autre part toute l'information nécessaire à une reconnaissance robuste n'est pas forcement contenue dans ces

quelques points clés, en effet beaucoup d'informations passent à la trappe lorsque l'image est compressée aux informations contenues à quelques endroits.

b-Méthode basées sur des apparences locales

Ici ce sont plusieurs vecteurs correspondant à des caractéristiques du visage qui sont utilisés en entrée. Ces méthodes sont à priori mieux taillées pour le problème à échantillon unique [10]. Tout d'abord car un set de plusieurs vecteurs de faible dimension au lieu d'un seul de grande dimension permet dès le début de s'attaquer à la malédiction de la dimensionnalité (curse of dimensionality). Ensuite le fait d'avoir plusieurs sous-vecteurs de caractéristiques permet l'utilisation d'un système de poids donnant priorité dans la décision finale aux sous-vecteurs identifiés comme étant les plus discriminatifs, ce qui améliore les performances [11]. Enfin, un nombre élevé de sous-vecteurs de caractéristiques peut augmenter la diversité des types de classifieurs utilisés par le biais d'équipe de classifieurs et ainsi améliorer les performances du classifieur global.

Variations	**Caractéristiques locales**	**Caractéristiques globales**
Petites variations	Pas sensible	Sensible
Grandes variations	Sensible	Très sensible
Illuminations	Pas sensible	Sensible
Expressions	Pas sensible	Sensible
Pose	Sensible	Très sensible
Bruit	Très sensible	Sensible
Occultations	Pas sensible	Très sensible

Tableau 1. 2.Comparaison des propriétés des caractéristiques locales et Des caractéristiques globales.

Les méthodes mentionnées ci-dessus ne considèrent pas explicitement les relations existantes entre les caractéristiques locales. Il est concevable que l'utilisation de cette information soit bénéfique pour le système de reconnaissance.

1.9.3 Méthodes hybrides

Les méthodes hybrides sont des approches qui combinent les caractéristiques globales et locales afin d'améliorer les performances de la reconnaissance de visage. En effet, les caractéristiques locales et les caractéristiques globales ont des propriétés tout à fait différentes. On peut espérer pouvoir exploiter leur complémentarité pour améliorer la classification.

1.10Méthode choisie

La méthode choisie est basée sur des caractéristiques d'apparence locales est une approche locale basée sur des modèles utilisent des connaissances à priori que l'on possède sur la morphologie d'image et s'appuient en général sur des points caractéristiques de celui-ci. En particulier, nous allons utiliser l'extraction des caractéristiques locales de l'histogramme de gradient orienté (HOG).

1.11L'extraction des caractéristiques

Cette étape représente le cœur du système de reconnaissance, on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision. Le choix de ces informations utiles revient à établir un modèle pour l'image, elles doivent être discriminantes et non redondantes. L'analyse est appelée représentation, modélisation ou extraction des caractéristiques.
L'efficacité de cette étape a une influence directe sur la performance du système de reconnaissance**.**

1.11.1Histogramme de gradient orienté (HOG)

Un histogramme de gradient orienté (HOG) est une caractéristique utilisée en vision par ordinateur. La technique calcule des histogrammes locaux de l'orientation du gradient sur une grille dense, c'est-à-dire sur des zones régulièrement réparties sur l'image. Il définit dans une région les proportions de pixels dont l'orientation du gradient appartient à un certain intervalle. Ces proportions caractérisent la forme présente dans cette région [12].

L'idée importante derrière le descripteur HOG est que l'apparence et la forme locale d'un objet dans une image peuvent être décrites par la distribution de l'intensité du gradient ou la direction des contours. Ceci peut être fait en divisant l'image en des régions adjacentes de petite taille, appelées cellules, et en calculant pour chaque cellule l'histogramme des directions du gradient ou des orientations des contours pour les pixels à l'intérieur de cette cellule. La combinaison des histogrammes forme alors le descripteur HOG. Pour de meilleurs résultats, les histogrammes locaux sont normalisés en contraste, en calculant une mesure de l'intensité sur des zones plus larges que les cellules, appelées des blocs, et en utilisant cette valeur pour normaliser toutes les cellules du bloc.

1.11.2Construction du descripteur HOG

a- Calcul du gradient

Une étape de prétraitement peut être effectuée avant le calcul du gradient, afin que les couleurs de l'image soient normalisées et une correction gamma correcte. Cette étape ne s'est finalement pas avérée nécessaire, la normalisation du descripteur lui-même s'avérant suffisante.

b- Construction de l'histogramme

La seconde étape est la création des histogrammes de l'orientation des gradients. Ceci est fait dans des cellules carrées de petite taille de (4x4 à 12x12) pixels. Chaque pixel de la cellule vote alors pour une classe de l'histogramme, en fonction de l'orientation du gradient à ce point. Le vote du pixel est pondéré par l'intensité du gradient en ce point.

c- Formation des blocs

Une étape importante est la normalisation des descripteurs afin d'éviter les disparités dues aux variations d'illumination. Cette étape introduit également de la redondance dans le descripteur. Pour cela, les auteurs regroupent plusieurs cellules dans un bloc, qui est l'unité sur laquelle est effectuée la normalisation. Les blocs se recouvrent, donc

une même cellule participe plusieurs fois au descripteur final, comme membre de blocs différents.

d- Normalisation des blocs

Selon l'éclairage ou le contraste entre le premier et l'arrière-plan, les valeurs des HOGs peuvent varier de façon importante, bien que la personne ait la même posture. Le modèle concerné de la base de données est alors moins proche et le descripteur est alors moins performant.

Afin d'uniformiser les vecteurs de caractéristiques, une normalisation du vecteur de caractéristiques **Vbloc**est réalisée. Soit **N** la taille du vecteur : [12]

$$\mathbf{V_{bloc}^{L1}(i) = \frac{V_{bloc}(i)}{\sum_{K=1}^{N} V_{bloc(k)}}} \qquad \textbf{(1.3)}$$

1.11.3Mesures de distance

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques issus du module d'extraction de caractéristiques d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence).

Nous allons utiliser la première catégorie de distances est constituée de distances Euclidiennes et sont définies à partir de la distance de Minkowski d'ordre **p** dans un espace euclidien [14].

La distance Euclidienne est une distance géométrique dans cet espace multidimensionnel. Il existe deux distances Euclidiennes :

a - Distance City-Block

Pour **p = 1** on obtient la distance City-Block :

$$\mathbf{L_1(x, y) = \sum_{i=1}^{n} |xi - yi|} \qquad \textbf{(1.4)}$$

b- Distance Euclidienne (L2)

Pour **p = 2** on obtient la distance euclidienne :

$$L_2(x, y)=\sqrt{\sum_{i=1}^{n}|xi - yi|2} \quad \textbf{(1.5)}$$

1.12Conclusion

L'état de l'art de la détection et de la reconnaissance d'un visage est très dense, nous y trouvons une multitude d'algorithmes que nous ne pouvons malheureusement tous citer. Cependant il est important de retenir que la détection d'un visage dans une image peut se faire selon une approche basée sur les connaissances acquises, une approche basée sur le templatematching, une approche basée sur l'apparence ou alors selon une approche basée des caractéristiques invariantes. Et concernant la reconnaissance de visage, elle s'effectue en utilisant soit une méthode globale, une méthode locale ou une méthode hybride.

Parfois, les systèmes de reconnaissance de visages présentent un manque de fiabilité au niveau de la décision à cause des erreurs commises lors de l'extraction et la classification des paramètres. Chaque technologie de capteur possède ses avantages et ses inconvénients. De ce fait, il apparaît comme impossible de faire une application fiable à 100%.

Chapitre2 : Reconnaissance automatique du locuteur

2.1 Introduction

Le terme générique « reconnaissance automatique du locuteur » est utilisé aussi bien pour définir l'identification et la reconnaissance du locuteur. La vérification consiste à accepter ou refuser l'identité proclamée par un locuteur, en se basant sur un modèle qui lui est associé. L'identification consiste en la reconnaissance d'un locuteur particulier parmi un ensemble fini de locuteurs possibles. Aussi bien la reconnaissance, que l'identification du locuteur se font en calculant un modèle stochastique sur la base de l'expression vocale du locuteur à reconnaître. Une fois calculé, ce modèle est comparé à des modèles prés entraînés sur la base de différentes phrases prononcées par les locuteurs.

La reconnaissance automatique du locuteur recherche des méthodes pour extraire les caractéristiques vocales propres à chaque individu. Ces caractéristiques servent à créer une signature vocale qui permette d'authentifier la voix de chacun.

Nous avons tous des timbres de voix différents. La voix de chaque personne dépend de caractéristiques à la fois anatomiques et comportementales.

La parole est le résultat de l'air faisant vibrer les cordes vocales et passant dans le conduit vocal constitué par la bouche et le nez. Si ces éléments anatomiques influencent la personnalité d'une voix, ils n'en fixent pas pour autant toutes les caractéristiques. Ainsi, une même personne ne parle pas tout le temps de la même façon. La voix change avec l'âge, l'humeur ou encore un rhume. En jargon scientifique, les variations de la voix d'une même personne sont appelées variabilité intra-locuteur.

En raison de ces aspects comportementaux, on parle de signature vocale, plutôt que d'empreinte [19].

Les applications potentielles des systèmes de reconnaissance de locuteur incluent le contrôle d'accès à distance de bases de données, les services d'information et de réservation à distance, les services bancaires à distance, etc. [20] La tendance actuelle montre une évolution vers l'exécution de diverses transactions en utilisant les téléphones mobiles.

2.2 Présentation du système de reconnaissance vocale

La reconnaissance automatique du Locuteur s'inscrit dans le domaine du traitement de la parole (figure 2.1). Elle consiste à reconnaitre l'identité d'une personne par l'analyse de sa voix.

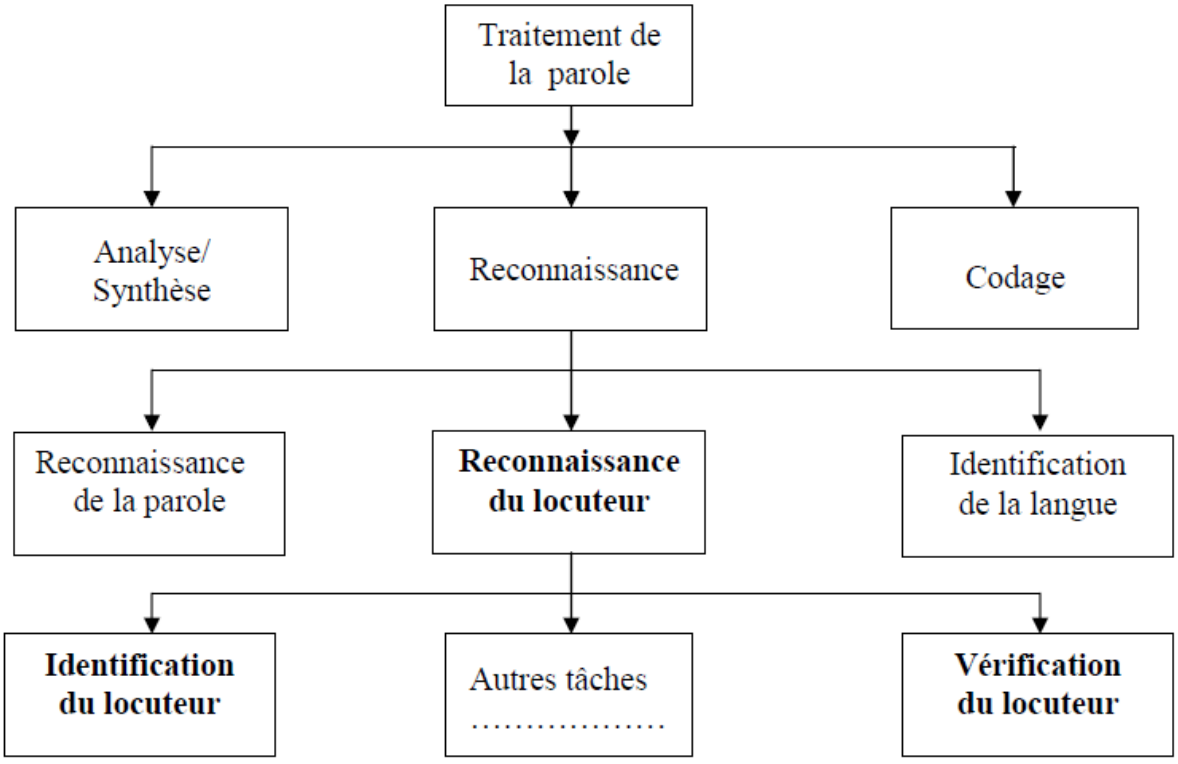

Figure 2. 1.Différentes taches du traitement de la parole [21].

2.2.1 Différentes tâches en reconnaissance du locuteur

La reconnaissance automatique de locuteur consiste à obtenir des renseignements concernant l'identité d'une personne à partir d'un enregistrement de sa voix. Pour qualifier précisément les différentes tâches entrant dans le cadre d'un système de reconnaissance automatique de locuteur, on distingue entre plusieurs tâches différentes :

2.2.1.1 La vérification *automatique* de locuteur (VAL)

Lorsqu'on cherche à décider si l'identité revendiquée par un locuteur est compatible avec sa voix. Dans ce type d'applications, il s'agit donc de trancher entre deux hypothèses, soit le locuteur est bien le locuteur autorisé, c'est à dire celui dont l'identité est revendiquée, soit nous avons affaire à un imposteur qui cherche à se faire passer pour un locuteur autorisé.

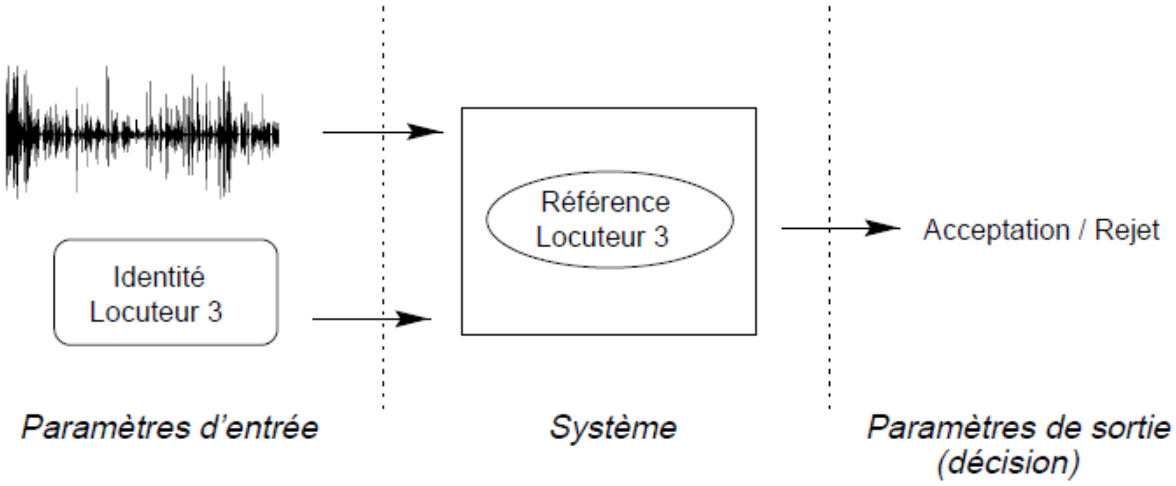

Figure 2. 2.La vérification automatique de locuteur.

2.2.1.2 L'identification automatique de locuteur (IAL)

Il s'agit de déterminer, parmi un ensemble de N locuteurs potentiels, à quel locuteur correspond un enregistrement vocal. En identification, la réponse apportée n'est pas de type binaire (acceptation ou rejet) comme pour la vérification, puisqu'il est nécessaire de distinguer un locuteur parmi un groupe. On distingue encore deux sous

problèmes d'identification selon que l'on est sûr ou non du fait que l'enregistrement provient bien d'un des membres du groupe de locuteurs :

- Si l'on a affaire à un ensemble fermé : le système IAL décide de l'identité la plus probable parmi les utilisateurs connus (dont il possède une référence). Ce mode de fonctionnement tend à considérer que seules des personnes référencées peuvent accéder au système. Un tel système ne doit alors être utilisé que dans un environnement au sein duquel tous les individus sont connus.

- Dans le cas d'un ensemble ouvert : le système IAL a la possibilité de rejeter le locuteur dont il teste les données si elles ne correspondent à aucune des identités répertoriées. Ce locuteur est alors considéré comme inconnu du système.

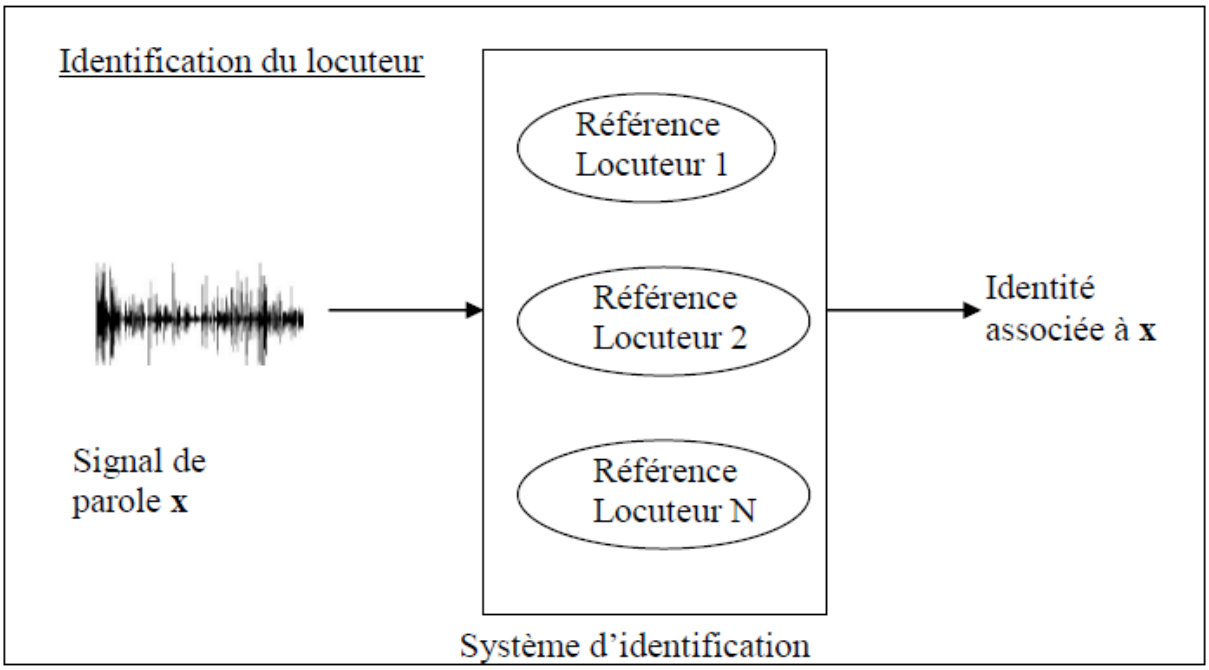

Figure 2. 3.Principe de base de l'identification du locuteur [22].

2.2.1.3 Détection de locuteur dans un flux multi-locuteurs.

Il s'agit d'une extension de la VAL à un test en environnement multi-locuteurs. Le principe est, à partir de l'enregistrement de référence d'un locuteur, de déterminer si ce locuteur est présent au sein d'un enregistrement multi-locuteurs.

2.2.1.4 Suivi de locuteur

Le suivi de locuteur consiste à trouver les limites des interventions du locuteur qu'on a recherché au sein du document multi locuteurs. Il s'agit donc de déterminer si ce locuteur intervient et si oui, quand.

2.2.1.5 Segmentation en locuteurs

C'est la détermination du nombre de locuteurs présents dans un document audio tout en délimitant leurs interventions. La complication de cette tache résulte du traitement de documents pour lesquels peu ou pas d'informations sont connues a priori. Notamment, pas d'information n'est disponible à la primitive concernant les locuteurs participant dans le document : ni leur nombre, ni leur identité, ni aucun échantillon de leur voix permettant d'avoir une référence. Toutes ces informations doivent être extraites du document étudié.

2.2.1.6 System dépendant et indépendant du texte

On classifie également les systèmes de reconnaissance et d'identification du locuteur en deux catégories :

a- Indépendant du contenu de la phrase prononcé ("text-independent") :

Les systèmes de RAL dits indépendants du texte si ne tiennent aucun compte du contenu linguistique du signal de parole.

b- Dépendant du texte et donc effectué sur la base d'un texte imposé ("text-dependent") : les systèmes dits dépendants du texte s'ils utilisent la connaissance de tout ou partie de ce contenu linguistique pour affiner la reconnaissance du locuteur.

2.3 Problématique

Les systèmes de RAL souffert des difficultés liées au domaine applicatif, comme l'utilisation des systèmes dans des conditions difficiles, les tentatives d'imposture, etc; la voix est beaucoup plus complexe qu'on ne peut percevoir par l'oreille. L'onde sonore varie non seulement avec les sons prononcés, mais également avec les locuteurs.

Ainsi l'obstacle majeur d'avoir une grande précision de la reconnaissance, est la grande variabilité des caractéristiques d'un signal vocal. Cette complexité du signal de parole provient de la combinaison de plusieurs facteurs :

2.3.1 Variabilité due au locuteur

Le signal parole varie pour un même individu parce que la voix d'une personne peut évoluer entre le début et la fin de la journée. Cette variabilité intra-locuteur est induite par l'évolution naturelle ou volontaire de la voix d'une personne.

2.3.2 Variabilité du canal

Une voix passant à travers un microphone, transmise par exemple par radio ou téléphone portable. Ces informations apparaissent le plus souvent sous la forme de déformations/dégradations du signal de parole.

2.3.3 Robustesse en environnements

Les systèmes de RAL doivent être robuste face au bruit ambiant et les environnements des canaux digitaux (téléphone, réseaux mobile, internet...). Un environnement calme ou bruité rend aussi plus ou moins facile la détection de la voix. Cette variabilité due au bruit environnant est difficilement prévisible et nécessite des traitements spécifiques pour être neutralisée.

2.4 Domaines d’application

Nos voix ne sont pas seulement un moyen de communiquer. Elles offrent également un moyen fiable de nous reconnaître, et font partie intégrante de notre identité. C'est la raison pour laquelle les banques et d'autres grandes entreprises se tournent aujourd'hui vers l'authentification vocale.

La voix humaine est unique. Elle est avec nous tout le temps contrairement à nos clés de voitures, et aux mots de passes ou codes PIN qu'on peut très souvent oublier. C'est à la fois cette sécurité et cette simplicité d'usage offerte par l'authentification biométrique vocale qui poussent les banques, les opérateurs de télécommunications

et autres grandes organisations à choisir ce mode d'authentification. On peut distinguer plusieurs domaines importants d'application dont notamment [20] :

2.4.1 Sécurisation des applications mobiles

Les grandes entreprises voient désormais leurs clients utiliser massivement les canaux mobiles pour prendre contact et effectuer les opérations courantes. C'est même devenu une attente forte des clients et des consommateurs. Mais la multiplication des applications et services en ligne fait qu'il devient difficile de gérer tous ces mots de passes, de forme et de tailles différentes. L'authentification vocale devient dès lors le mode d'authentification mobile idéal. Il suffit simplement de donner une simple phrase clé à prononcer à un client pour vérifier son identité.

En plus d'éliminer la frustration née des mots de passe difficiles à mémoriser ou à saisir, le 'login vocal' réinvente véritablement l'authentification mobile. Le mobile devenant de plus en plus le point de contact principal entre un consommateur et un fournisseur de services, améliorer l'expérience utilisateur et la sécurité deviennent une priorité.

2.4.2 Sécurisation des transactions à risque par carte de crédit

La reconnaissance de locuteur constitue aussi une solution sûre et pratique pour vérifier les transactions à risque par carte de crédit (par exemple celles en dehors des habitudes de consommation du client ou de son emplacement géographique habituel). Quand une opération à risque est détectée, une demande de vérification de la transaction peut être envoyée au titulaire de la carte de crédit, via un appel sortant automatique, sur son téléphone portable. Le détenteur est alors invité à prononce une phrase clé : "J'autorise cette transaction par ma signature vocale".

2.4.3 Paiement en ligne

La reconnaissance de la voix peut être utilisée pour sécuriser des paiements en ligne, typiquement des paiements à risque tels que le premier paiement en ligne sur un site d'e-commerce, par exemple le transfert de l'argent ou des opérations importantes. Lorsque ces opérations sont effectuées, un appel sortant automatique est émis vers le

téléphone portable du titulaire du compte effectuant l'opération. Si cette opération est valide, l'utilisateur est invité à confirmer le paiement de la même façon qu'il peut confirmer l'achat par carte de crédit.

2.4.4 Aide aux handicapés

La reconnaissance de locuteur est très utile dans ce cas, elle offre la possibilité de saisir les données à la voix, commandes vocales (ouverture porte, contrôle des équipements au domicile).

2.5 Les méthodes de reconnaissance du locuteur

Un système de reconnaissance (identification ou vérification) comporte plusieurs composantes : un module d'extraction de paramètres, un bloc d'appariement, un module de normalisation des scores d'appariement et un module de décision.

Nous présentons par la suite les différentes approches et techniques utilisées en extraction de paramètres ainsi modélisation et décision.

2.5.1 Extraction de paramètres

Le module d'extraction de paramètres transforme un signal de parole en une séquence de vecteurs acoustiques utiles à la reconnaissance. Ce module comporte différents sous-modules à savoir, la paramétrisation, la segmentation parole / non parole (Speech/non-speech) et des prétraitements.

Pour cela, il existe déjà plusieurs techniques :

2.5.1.1 La transformée de Fourier discrète :

L'analyse spectrale se fait à l'aide de la transformée de Fourier. Le signal parole est un signal non stationnaire à long terme mes présumé stationnaire pour une durée allant de 10 à 30 ms. Pour cela on applique le Transformé de Fourier à court terme (FFT- Fast Fourier Transforme), où il est nécessaire d'effectuer préalablement un fenêtrage, s'appelle des trames allant de 20 à 30 ms avec un recouvrement. Avec X (n) le spectre du signal numérique x (k) :

$$X(n)=\sum_{k=0}^{N-1}x(k)\times e^{-j\pi\frac{nk}{N}} \quad \textbf{(2.1)}$$

2.5.1.2 Coefficient MFCC

Le signal acoustique contient de différentes sortes de renseignements sur le locuteur. La paramétrisation MFCC (Mel FrequencyCepstral Coefficients) est basée sur la perception humaine de son, sur l'évidence connue que les renseignements portés par les composantes de la fréquence basse du signal de parole sont plus importants phonétiquement pour les humains que les composantes à haute fréquence.

Nous donnerons brièvement les étapes de leur calcul :

a) Fenêtrage du signal avec la fenêtre de Hamming.

b) Transformée de Fourier.

c) Filtrage par banc de filtres triangulaires espacés selon l'´echelle Mel.

d) Transformée en Cosinus Discrète.

e) Centrage et mise à 1 de la variance du vecteur calculé.

2.5.1.3 Les paramètres AR

La modélisation d'un signal x (k) consiste à lui associer un filtre linéaire qui soumit à une excitation particulière reproduit ce signal le plus fidèlement possible. L'objectif essentiel de la modélisation AR d'un signal est de permettre sa description par un ensemble très limité de paramètres :

$$x(k)+a_1x(k-1)+....+a_px(k-p)=e(k) \quad \textbf{(2.2)}$$

Les coefficients de la combinaison linéaire sont trouvés de façon à minimiser l'erreur, Cette modélisation correspond à un modèle tout pôle ; Soit H (z) sa fonction de transfert :

$$H(z)=\frac{1}{A(z)} \quad \text{Où} \quad A(z)=1-\sum_{i=1}^{p} a_i z^{-1} \qquad \textbf{(2.3)}$$

2.5.1.4 Les coefficients PLP (perceptuel linear predictive)

La méthode de coefficients de prédiction à base de notions psycho acoustiques connue sous le nom PLP, est une méthode inspirée du principe de prédiction linéaire (LPC). Elle combine ce principe à une représentation du signal qui suit l'échelle humaine de l'audition. Le principe de la méthode dont une analyse spectrale est effectuée au signal parole afin d'obtenir un spectre suivant une échelle d'audition. Ce spectre est ensuite modifié par une interpolation et une transformée de Fourier inverse, le signal obtenu étant passé dans un filtre pour réduire la dimension du spectre et augmenter la résolution fréquentielle.

2.5.2 Modélisation

Les techniques de modélisation et extraction de paramètres sont les parties principales d'un système de reconnaissance du locuteur.

On peut distinguer quatre grandes approches pour la construction des modèles de locuteur : l'approche vectorielle, statistique, prédictive et connexionniste.

2.5.2.1 Approche vectorielle

Dans l'approche vectorielle, les vecteurs de paramétriques d'apprentissage et de test sont comparés, sous l'hypothèse que les vecteurs d'une des séquences sont une réalisation imparfaite des vecteurs de l'autre séquence. La distorsion entre les deux séquences représente leur degré de similarité. L'approche vectorielle compte deux grandes techniques : la programmation dynamique et la quantification vectorielle.

a-La programmation dynamique

La programmation dynamique (Dynamic Time Warping : DTW) consiste à aligner temporellement une séquence de vecteurs de paramètres de test avec une séquence de vecteurs d'apprentissage. Dans ce cas, le modèle de locuteur est tout simplement l'ensemble des vecteurs de paramètres obtenus après paramétrisation des signaux d'apprentissage.

Une distance est calculée entre vecteurs d'apprentissage et de test et moyennée sur l'ensemble de la séquence.

La programmation dynamique est utilisée exclusivement en mode dépendant du texte, c'est une approche très rapide et fournie des résultats relativement bonne, mais elle est très sensible à la qualité d'alignement et notamment au choix du point de départ.

b-La quantification vectorielle

La quantification vectorielle (VectorQuantization : VQ) repose sur un partitionnement de l'espace acoustique en sous-espaces. Chaque sous-espace est associé à leur vecteur centroïde (un vecteur de paramètres représentant l'ensemble des vecteurs composant le sous-espace). Dans ces conditions, un modèle de locuteur est composé d'un ensemble de vecteurs centroïdes, appelé dictionnaire de quantification (codebook).

Lors de la phase de reconnaissance, une distance est calculée entre un vecteur de test et chaque vecteur centroïde du dictionnaire. La distance minimale est assignée au vecteur de test. La distance d'une séquence de vecteurs de test est obtenue par moyenne des distances minimales attribuées à chacun des vecteurs de test.

La quantification vectorielle s'applique en mode dépendant ou indépendant du texte. La rapidité et les performances de cette technique dépendent fortement de la taille du dictionnaire : plus la taille du dictionnaire augmente, meilleures sont les performances sinon, le processus devient plus lent.

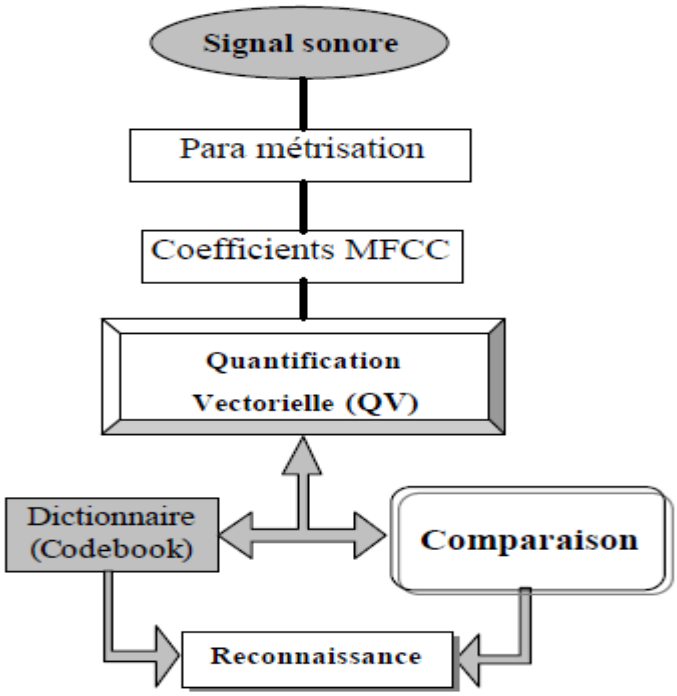

Figure 2. 4.Reconnaissance du locuteur à base de Quantification vectorielle (QV) [24].

2.5.2.2 Approche statistique

Les techniques statistiques considèrent le locuteur comme étant une source probabiliste et le modélisent par une densité de probabilité connue. La phase d'apprentissage consiste à estimer les paramètres de la fonction de densité de probabilité. La décision est prise en calculant la vraisemblance des données par rapport au modèle appris préalablement. Les modèles de Markov cachés HMM ont été utilisés dans des applications dépendantes du texte de reconnaissance automatique du locuteur tandis que les Modèles de Mélange de lois Gaussiennes GMM et les machines à vecteurs de support SVM largement utilisés en indépendant du texte ainsi dans des applications de vérification du locuteur.

2.5.2.3 Approche prédictive

Les modèles prédictifs reposent sur le principe qu'une trame d'un signal de parole peut être générée à partir des trames précédentes du signal. Un locuteur donné est représenté par une fonction de prédiction estimée sur ses données d'apprentissage.

Deux stratégies peuvent être ensuite adoptées pour la reconnaissance : soit calculer une erreur de prédiction comme mesure de similarité, entre les trames et les trames réellement observées ; soit comparer la fonction de prédiction du locuteur concerné avec une nouvelle fonction de prédiction estimée cette fois ci sur les nouvelles données, selon une mesure de distance donnée.

2.5.2.4 Approche connexionniste

L'approche connexionniste repose sur la discrimination entre locuteurs. Elle consiste à fournir à un réseau de neurones un ensemble de signaux de parole issus d'une population de locuteurs afin que ce dernier apprenne comment discriminer un locuteur des autres. L'approche connexionniste se résume, par conséquent, à une tâche de classification. Un modèle se présente sous la forme d'un ou plusieurs réseaux de neurones pour lequel la séquence de vecteurs d'apprentissage du locuteur concerné ainsi que celles des autres locuteurs du système sont fournies en entrée.

2.5.3 Décision et mesures de performances

La stratégie mise en jeu dans cette partie dépend essentiellement des deux processus : la vérification et l'identification automatique de locuteur.

a- Pour l'identification du locuteur on a la relation suivante :

$$I_c = \frac{Nombre\ de\ tests\ \ correctement\ \ identifiés}{Nombre\ total\ \ de\ \ tentatives}$$

$$I_i = \frac{Nombre\ \ de\ \ tests\ \ mal\ \ identifiés}{Nombre\ \ total\ \ de\ \ tentatives} \qquad \textbf{(2.4)}$$

$$\text{Avec : } I_c + I_i = 100\%$$

Les performances du système d'identification sont données en termes de taux D'identification correcte I_c ou incorrecte I_i.

b- La vérification du locuteur. C'est une décision en tout ou rien :

$$FR = \frac{Nombre\ \ de\ \ tentatives\ \ d'abonnésrejetées}{Nombre\ \ total\ \ de\ \ tentatives\ \ d'abonnés}$$

$$FA = \frac{Nombre\ \ de\ \ tentatives\ \ d'imposteur\ acceptées}{Nombre\ \ total\ \ de\ \ tentatives\ \ d'imposteurs} \qquad \textbf{(2.5)}$$

Les performances de la vérification du locuteur sont données en termes des faux rejets FR et de fausses acceptations FA.

2.5.3.1 Distances et mesures de distance

Il est possible d'utiliser toutes les distances classiques, les distances de Minkovski, parmi lesquelles la distance euclidienne, et la distance de Mahalanobis qui normalise les coefficients par leur matrice de covariance. Des distances spécifiques aux espaces de représentation de parole existent aussi, comme les distances cepstrales pondérées, la mesure d'Itakura pour comparer les modèles autorégressifs, ou encore la distance par appariement de Pics Spectraux.

2.6 Méthodes choisies et les raisons du choix

Nous cherchons dans ce travail à adapter les coefficients cepstraux qui peuvent donner plus de fidélité au locuteur.

Une extension possible des cepstres est leur passage dans un espace fréquentiel non linéaire proche de l'audition humaine. Il est ainsi possible de modifier la procédure de calcul précédente pour que les coefficients obtenus soient répartis selon une échelle Mel. Une telle procédure, proposée dans [25], permet d'obtenir des coefficients cepstraux à échelle Mel, *MelFrequencyCepstral Coefficients*, *MFCC*.

Nous voyons deux avantages à l'emploi de la méthode des MFCC qui permet, en outre, d'obtenir une information synthétique sur le signal de parole de meilleure qualité que la transformée de Fourier tout en utilisant un espace de représentation plus restreint.

La première qualité de la méthode MFCC est sa résistance reconnue au bruit. La deuxième qualité majeure de la méthode MFCC est sa plausibilité biologique puisqu'elle utilise une échelle psychoacoustique des fréquences similaires à celle de l'oreille interne.

Nous considérons ici dans ce travail la modélisation par la méthode de quantification vectorielle. La méthode de QV s'est avérée robuste en termes de reconnaissance.

Pour le cas de la mesure spectrale, la distance la plus utilisée est La distance cepstrale euclidienne. Donc dans l'étape de décision, la distance euclidienne sera utilisée.

À travers la recherche s'est avéré être que l'étape d'identification est l'étape essentielle dans la reconnaissance du locuteur, nous sommes devenus donc plus intéressés à identifier le locuteur que la vérification dans ce travail.

2.6.1 Extraction des vecteurs MFCC

La méthode MFCC (Mel FrequencyCepstral Coefficients) est une méthode d'extraction des paramètres selon l'échelle de Mel. En effet, la perception de la parole par le système auditif humain est fondée sur une échelle fréquentielle semblable à l'échelle de Mel [29]. Cette échelle est linéaire aux basses fréquences et logarithmique en hautes fréquences et elle est donnée par l'équation suivante:

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \qquad \textbf{(2.6)}$$

Bien qu'il existe plusieurs méthodes d'extraction de vecteurs acoustiques, lareprésentation basée sur une échelle non linéaire nommée l'échelle de Mel (c.-à-d.MFCC) est la plus utilisée. En outre, la principale caractéristique de cette échelle est sa simulation du mécanisme perceptuel non linéaire de l'oreille humaine.

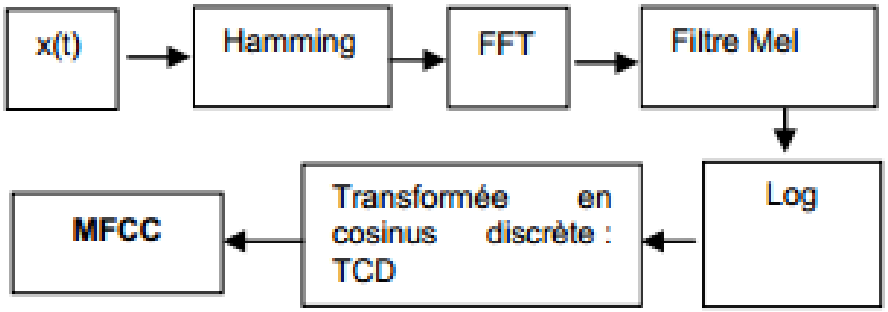

Figure 2. 5.Calcul des coefficients MFCC avec une échelle Mel.

Dans ce qui suit, nous décrirons les étapes principales du processus d'extraction des vecteurs acoustiques de type MFCC :

a-Fenêtrage : consiste en premier temps à découper le signal en trames chevauchées de faible durée où le signal est considéré comme quasi stationnaire. Ensuite, chaque trame est multipliée par une fenêtre temporelle d'analyse qui peut prendre plusieurs formes, uniformes, triangulaires, gaussiennes, etc. Toutefois, la fenêtre de Hamming reste la plus utilisée dans le domaine du traitement de la parole. En fait, l'objectif principal de l'étape du fenêtrage du signal est d'atténuer les discontinuités du signal au bout de ces trames tout en réduisant le signal à zéro autour de ces extrémités. Le choix de la taille de la fenêtre d'analyse représente toujours un dilemme du fait qu'une fenêtre de très courte durée assure l'hypothèse de la quasi-stationnarité du signal. Cependant elle ne contient pas suffisamment d'échantillons pour assurer une bonne estimation des paramètres du signal.

b– Transformée de Fourier rapide (Fast Fourier Transform, FFT) : est un algorithme conçu pour calculer rapidement la transformée de Fourier discrète. La FFT sera appliquée pour chaque fenêtre d'analyse (c.-à-d. trame fenêtrée) afin de réaliser le passage du signal du domaine temporel au domaine fréquentiel.

c– Filtrage selon l'échelle de Mel : l'échelle de la perception fréquentielle de l'oreille humaine n'est pas linéaire. Par conséquent, une filtration du signal vocal par une banque de filtres positionnés selon une échelle similaire à la nôtre peut alléger largement la complexité du traitement. L'échelle de Mel est construite à partir d'une série de filtres passe-bandes de formes triangulaires (voir Figure 2.6) positionnés d'une

façon linéaire, pour les basses fréquences (< 1000 Hz) et logarithmique pour les hautes fréquences. Cette échelle reproduit la sélectivité de l'oreille qui diminue avec l'accroissement de la fréquence.

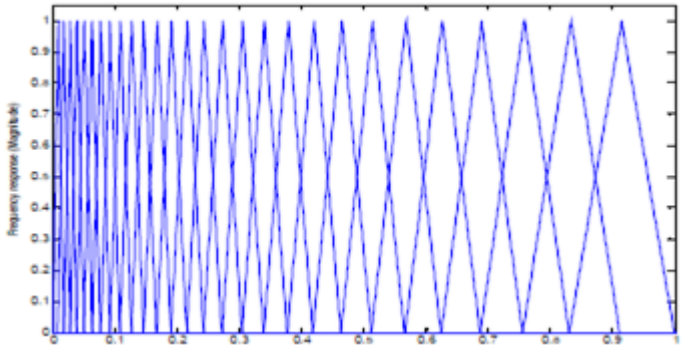

Figure 2. 6.Banc de filtres dans l'échelle de Mel-fréquence.

d-Transformée en cosinus discrète: cette transformée consiste à multiplier les logarithmes des réponses en énergie des filtres de Mel par des fonctions sinusoïdes de différentes fréquences. L'objectif est de décorréler ces valeurs d'énergie pour constituer par la suite les coefficients (éléments) de notre vecteur MFCC final.

2.6.2 Quantification vectorielle (QV)

La quantification vectorielle décompose l'espace acoustique d'un locuteur donné X, en un ensemble de M sous-espaces représentés par leur vecteurs centroïdes C = {c1, c2,....cM }. Ces vecteurs centroïdes forment un dictionnaire (de taille M) qui modélise ce locuteur, et sont calculés en minimisant l'erreur de quantification moyenne (distorsion) induite par le dictionnaire sur les données d'apprentissage du locuteur {x1, x2,…xM}:

$$D(X,C) = \frac{1}{T}\sum_{t=1}^{T} \min d\left(x_t, c_m\right) \quad , \qquad 1 \le m \le M \tag{2.7}$$

Où $d\ (x_t, c_m)$ est une mesure de distance au sens d'une certaine métrique liée à la paramétrisation. L'apprentissage vise à réduire l'erreur de quantification. On peut mieux représenter le locuteur en augmentant la taille du dictionnaire, mais le système sera moins rapide et plus demandeur de mémoire. Il faut trouver donc un bon

compromis. La construction du dictionnaire peut être faite en utilisant par exemple l'algorithme LBG [27].

Dans l'illustration, apparaissent seulement deux locuteurs et deux dimensions de l'espace acoustique. Les cercles désignent les vecteurs acoustiques du locuteur 1 alors que les triangles sont du locuteur 2. Dans la phase de formation (apprentissage), à l'aide de l'algorithme de clustering décrite par la suite, un livre de codes (codebook) VQ spécifique est généré pour chaque locuteur connu par ses vecteurs acoustiques de la formation de cluster. Les mots décode de résultat (le centre de gravité (centroïde)) sont indiqués dans la Figure 2.7 de cercles noirs et des triangles noirs pour locuteur 1 et 2, respectivement. La distance entre un vecteur et le mot de code le plus proche d'un livre de codes est appelée une VQ-distorsion. Dans la phase de reconnaissance, on forme le VQ-distorsion » à l'aide de chaque codebook d'énoncé d'entrée d'une voix inconnue. Le locuteur correspondant au codebook VQ avec plus petite distorsion totale est identifié comme le locuteur de l'énoncé d'entrée.

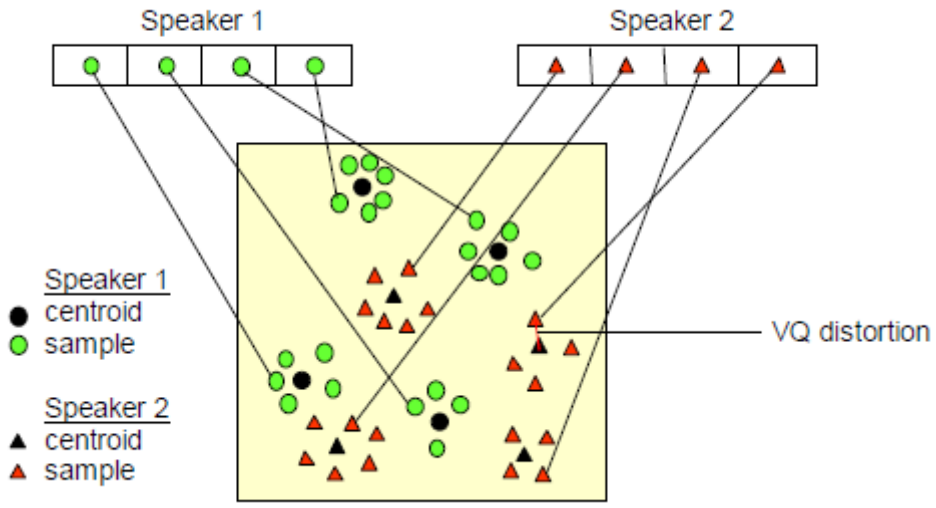

Figure 2. 7.Diagramme conceptuel illustrant un dictionnaire de codes (codebook) pour le (VQ).

2.6.2.1 Cluster les vecteurs d'apprentissage

Après l'étape d'apprentissage, les vecteurs acoustiques extraits fournissent un ensemble de vecteurs d'apprentissage. La prochaine étape importante est de construire un dictionnaire VQ spécifique de chaque locuteur en utilisant ces vecteurs d'apprentissage par l'algorithme de LBG [26]. Pour cluster un ensemble de vecteurs "L" des vecteurs d'apprentissage dans un ensemble de "M" vecteurs de dictionnaire

(codebook). L'algorithme est officiellement mis en œuvre par la procédure récursive suivante :

a. Concevoir un livre de codes 1-vecteur ; c'est le centre de gravité (centroïde) de l'ensemble de vecteurs d'apprentissage (par conséquent, aucune itération n'est nécessaire ici).

b. Doubler la taille de codebook en divisant chaque codebook en cours n y selon la règle:

$$y_n^+ = y_n(1+\varepsilon)$$

$$y_n^- = y_n(1-\varepsilon) \qquad \textbf{(2.8)}$$

Où 'n' varie de 1 à la taille actuelle de la table de codage (codebook), et ε est un paramètre de partage. (Généralement on choisit ε=0.01).

c. Recherche le proche voisin : pour chaque vecteur d'apprentissage, trouver le mot de code (codeword) dans le livre de code (codebook) actuel qui est la plus proche (en termes de mesure de similarité), et attribuer ce vecteur à la cellule correspondante (associé avec le mot de code le plus proche).

d. Le centroïde à mettre à jour : mettre à jour le mot de code (codeword) dans chaque cellule en utilisant les centroides des vecteurs d'apprentissage affectés à cette cellule.

e. Itération 1 : répétez les étapes c et d jusqu'à ce que la distance moyenne soit inférieure à un seuil prédéterminé.

f. Itération 2 : répéter les étapes b, c et d jusqu'à une taille de bibliothèque de codes de M soit conçu.

Intuitivement, l'algorithme LBG conçoit un codebook de M-vecteurs. On commence d'abord par la conception d'un vecteur de codebook. Puis on utilise une technique de séparation sur les codeword pour initialiser la recherche de 2-vecteurs de codebook, et le processus se continue jusqu'à ce que le fractionnement de la table de codage M-vecteurs désiré soit obtenu. La figure2.8représente le diagramme avec les étapes

détaillées de l'algorithme LBG. "Vecteurs de cluster" est la plus proche voisine procédure de recherche qui attribue chaque vecteur d'apprentissage à un groupe associé au codeword le plus proche [28]. Trouvez les centroïdes est la procédure essentielle. "Calculer D (distorsion)" résume les distances de tous les vecteurs d'apprentissage à la recherche du plus proche voisin afin de déterminer si la procédure a convergé.

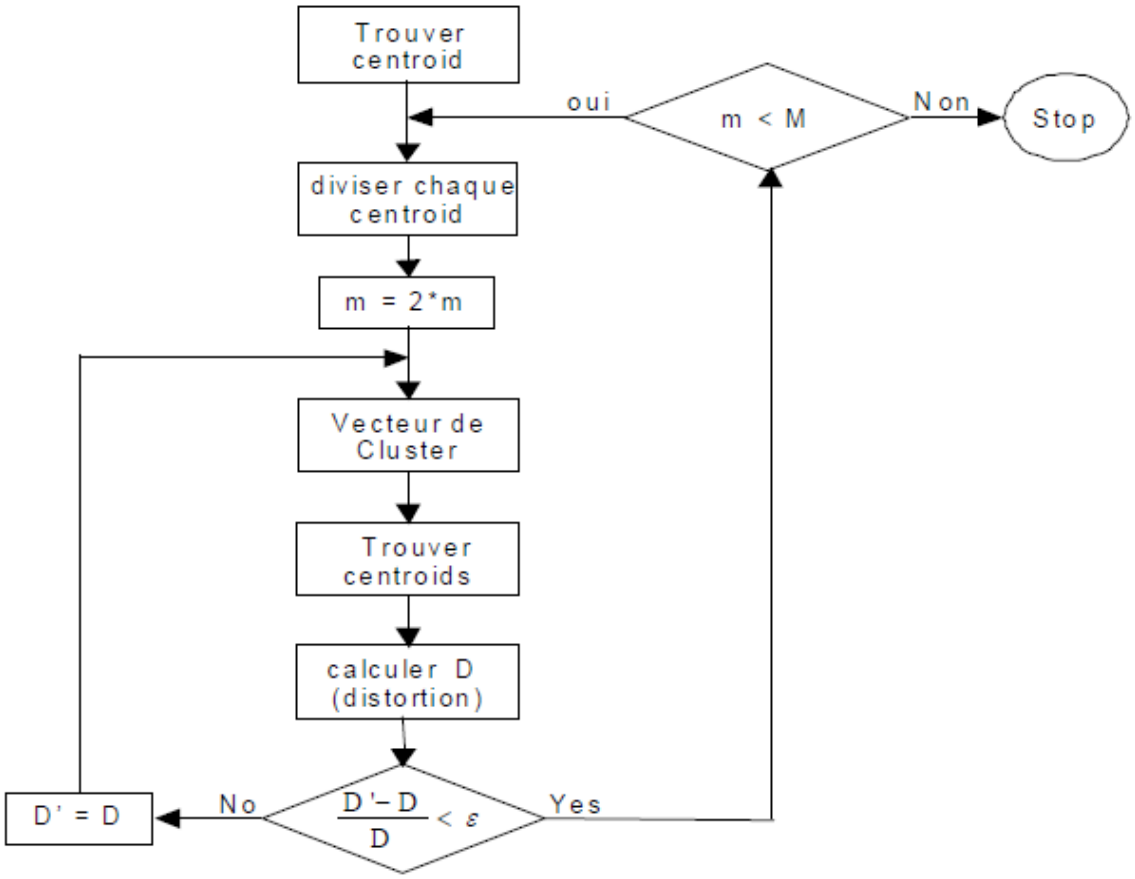

Figure 2. 8.Diagramme de LBG [28].

2.6.3 Distance euclidienne

Dans notre travail, nous avons utilisé la méthode de classificateur de Distance euclidienned_2 (DE) :

$$d_n(x,y) = \left(\sum_{k=1}^{p} |x_k - y_k|^n \right)^{\frac{1}{n}} \quad \textbf{(2.9)}$$

Avec **n=2**

La distance euclidienne classificatrice (DE) a l'avantage de la simplicité et la rapidité de calcul. Le classement se fait par calcul de la distance minimale à fin de décider lequel des locuteurs sur tout l'ensemble d'apprentissage et les plus susceptibles d'être le locuteur de test.

2.7 Conclusion

Dans ce chapitre, nous avons présenté de façon générale l'état de l'art d'un système de reconnaissance automatique de locuteur, nous avons présenté, également, la structure générale d'un système RAL et ses composants modulaires. Pour chaque module, nous avons décrit les différentes techniques utilisées en citant leurs avantages et leurs faiblesses, et nous avons terminé par la présentation des domaines d'application de cette discipline.

Un système de reconnaissance automatique du locuteur, quelle que soit la tâche considérée, se résume en trois étapes principales : l'analyse acoustique du signal parole, la modélisation du locuteur et la décision soit une décision et vérification. Egalement, tout système de RAL dépend de la technique d'extraction de paramètres utilisé, modélisation, décision et ainsi la phase de prétraitement.

}

Chapitre 3 : Fusion de l'information

3.1 Introduction

La fusion de l'information est un sujet relativement ancien qui trouve ses origines à partir du moment où les chercheurs ont fait leurs premières tentatives d'imitation de l'intelligence humaine.

Le concept de fusion d'informations est naturel. Par exemple, les êtres humains peuvent combiner les perceptions des différents sens pour construire une image mentale unifiée de l'environnement. Il a toutefois fallu attendre l'émergence de l'informatique pour transposer aux capteurs artificiels la capacité naturelle des êtres vivants à fusionner des informations.

La fusion d'informations est apparue afin de gérer des quantités très importantes de données multi sources. Plusieurs sens sont donnés à la fusion d'informations, nous reprenons ici la définition proposée par (Bloch 2003) : [34] La fusion d'informations consiste à combiner des informations issues de plusieurs sources afin d'aider à la prise de décision.

Actuellement, la profusion des capteurs, l'amélioration des processeurs et les progrès en modélisation formelle tendent à favoriser un usage de plus en plus répandu de la fusion d'informations. [35] La fusion d'informations peut alors se définir comme la combinaison d'informations (souvent imparfaites et hétérogènes) afin d'obtenir une information globale plus complète, de meilleure qualité, et permettant de mieux décider et agir.

Une formalisation possible de la fusion d'information introduit trois niveaux conceptuels correspondant à trois types d'information : celui des données (ou bas niveau), celui des caractéristiques (ou fusion de niveau intermédiaire) et celui des décisions (ou fusion de haut niveau).

Le choix du niveau de fusion doit se faire en fonction des données disponibles et de l'architecture de la fusion retenue qui sont liées à l'application recherchée.

3.2 Niveau de fusion

La combinaison de plusieurs systèmes peut se faire à trois niveaux différents : au niveau des données, au niveau des caractéristiques extraites, au niveau des décisions.

3.2.1 La fusion de données

C'est le niveau conceptuel le plus bas. Elle consiste essentiellement à marier des informations de bas niveau comme par exemple des primitives, dans le but de rendre l'information moins bruitée que celle obtenue avec une seule source d'information.

3.2.2 Fusion au niveau caractéristique

Chaque processus produit une série des caractéristiques. Le processus de fusion combine ces collections de caractéristiques en un unique ensemble ou vecteur de caractéristiques.

3.2.3 Fusion au niveau de décision

La fusion au niveau de décision est souvent utilisée pour sa simplicité. En effet, chaque système fournit une décision binaire sous la forme OUI ou NON que l'on peut représenter par 0 et 1, et le système de fusion de décisions consiste à prendre une décision finale en fonction de cette série de 0 et de 1.

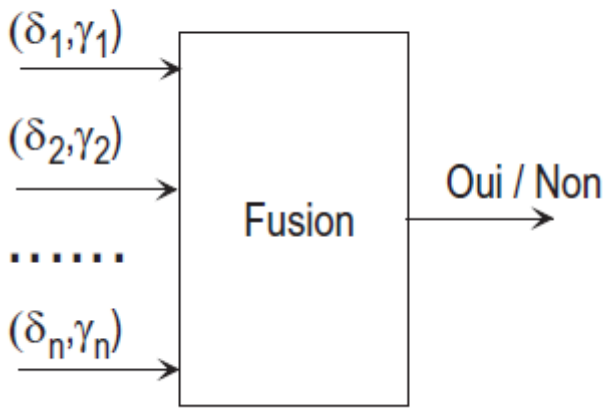

Figure 3. 1.Fusion de décision, avec sortie binaire (oui/non) [31].

3.3 Fusion a base des methodes non parametriques : [44]

Ces méthodes n'utilisent que des informations du premier ordre (sorties declassificateurs). Elles sont faciles à implémenter et ne nécessitent pas de phase d'apprentissage.Cependant, le point faible de ces méthodes est qu'elles traitent les classificateurs de manières équivalente ce qui ne permet pas de tenir de leur capacité individuelle.

3.3.1 Fusion endécision

La combinaison d'un ensemble de classificateur de type classe est souvent basée sur le principe du vote. Ces méthodes consistent à interpréter chaque sortie d'un classificateur comme un vote pour l'une des classes possibles. La classe ayant un nombre de votes supérieur à un seuil préfixé est retenue comme décision finale. Elles sont les plus simples à mettre en œuvre car les votes ne sont pas pondérés et chaque classe reçoit autant de votes qu'il a de classificateurs à combiner.

Les méthodes de vote peuvent pratiquement toutes être dérivées de la règle avec seuil exprimée par :

$$E(x) = \begin{cases} C_i \quad si \sum_{j=1}^{L} e_{i,j} = \max \sum_{j=1}^{L} e_{t,j} \geq \lambda L \\ \text{rejet} \quad \text{sinon} \end{cases} \qquad \textbf{(3.1)}$$

λ correspond à la proportion de classificateur devant répondre par la même classe pour que celle-ci soit retenue comme résultat de la combinaison. Ainsi, pour λ = 0, il s'agit du vote à la pluralité où la classe qui reçoit le plus de votes est choisie comme classe finale.On parle de majorité notoire dans le cas où pour être désignée comme réponse finale, la classe majoritaire, en plus elle doit se distinguer de la deuxième classe d'une différence supérieure à un seuil fixé.Le principal inconvénient de ces méthodes est que toutes les classes possèdent lemême vote ce qui sera considéré comme un conflit. Pour remédier un ce problème, onchoisit d'utiliser les classificateurs de type rang en observant non seulement les premières réponses, mais les K premières classes ordonnées par rang et on les comptabilise dans le vote. [43]

3.4 Fusion a base des methodesparametriques :

Comparées aux autres méthodes, les méthodes de fusion paramétriques sont pluscomplexes à mettre en œuvre. Elles utilisent des paramètres supplémentaires calculéspendant la phase d'apprentissage. La performance de ces méthodes dépend alors de la bonne estimation des paramètres lors de l'entraînement.

3.4.1 Fusion endécision :

On a principalement le vote pondéré, tel que chaque vote du classificateur est pondéré par sa fiabilité $W_j$et on a :

$$E(x) = \begin{cases} C_i \quad si \sum_{j=1}^{L} W_j \, e_{i,j} = \max \sum_{j=1}^{L} W_j \, e_{t,j} \geq \lambda L \\ \text{rejet} \quad \text{sinon} \end{cases} \qquad \textbf{(3.2)}$$

En général, W_j = taux de reconnaissance / 100-taux de rejet est calculé pendant untest d'apprentissage où on simule des reconnaissances pour évaluer la fiabilité de chaque système séparément. Ainsi, la forme d'entrée est attribuée à la classe pour laquelle la somme des votes, qui sont pondérés par la fiabilité estimée de chacun des experts, est la plus élevée.

3.5 Méthodes de fusion d'informations

Nous présentons dans cette section l'état de l'art des méthodes de fusion d'information, de haut niveau. On trouve dans la littérature quatre principales méthodes de fusion d'information : la fusion par le vote, par la théorie des probabilités, par la théorie des possibilités et par la théorie des fonctions de croyance.

3.5.1 Principe du vote

La méthode de fusion d'information la plus simple à mettre en œuvre et la plus naturelle est le vote. À partir de décisions, souvent binaires, émises pour chaque classe par plusieurs sources, le vote a pour principe de choisir la classe qui possède le plus d'occurrences. L'un des moyens de procéder est d'utiliser plusieurs classifieurs, chaque classifieur permettant l'obtention d'une décision binaire pour chacune des classes.

Lorsque plusieurs sources d'information sont disponibles, un classifieur peut être construit pour chacune des sources. Un vote à partir des décisions émises par chacun des classifieurs permet de choisir une classe en tenant compte des différentes sources.

Les méthodes les plus utilisées sont des méthodes à base de votes telles que le OR (si un système a décidé 1 alors OUI), le AND (si tous les systèmes ont décidé 1 alors OUI) ou le vote à la majorité (si la majorité des systèmes ont décidé 1 alors OUI).

3.5.2 Fusion par approche probabiliste

L'approche probabiliste, permet de modéliser l'incertitude en affectant à une donnée une probabilité d'appartenance à chaque classe. Ayant une unique mesure pour décrire la probabilité d'appartenance d'une donnée à une classe, l'imprécision ne peut pas être distinguée de l'incertitude. Il y a donc une confusion entre les incertitudes dues au bruit, et les imprécisions dues à un manque de connaissance. Confondues dans l'approche probabiliste, ces deux imperfections sont indûment nommées incertitudes. Cette théorie a cependant pour l'avantage d'avoir été beaucoup étudiée et de posséder un formalisme précis régie par le théorème de Bayes [36]. Afin de combiner les probabilités estimées à partir de plusieurs sources d'information, on recourt à l'utilisation de la règle de Bayes :

$$P(X|S_1,\ldots,S_N) = \frac{P(X)P(S_1,\ldots,S_N|X)}{P(S_1,\ldots,S_N)} \quad \textbf{(3.3)}$$

Où X et $S_1,\ldots,S_N$ sont respectivement les classes et l'ensemble des sources d'information à fusionner. Sous l'hypothèse d'indépendance statistique entre les sources, ce modèle de combinaison peut être simplifié en un simple produit des probabilités. Notons que dans la plupart des applications, l'indépendance cognitive est admise. $

3.5.3 Fusion par approche possibiliste

La théorie des possibilités [37] permet de tenir compte de l'imprécision des données ainsi que de l'incertitude à partir de deux fonctions de possibilité et de nécessité. Ces deux fonctions sont obtenues à partir des distributions de possibilités définies sur $D = \{C_1,\ldots,C_n\}$ par :

$$\pi : D \to [0,1], \sup_{x\in D} \pi(x) = 1. \quad \textbf{(3.4)}$$

Ces distributions donnent le degré d'appartenance au domaine D, qui n'est autre qu'un opérateur flou.

La théorie des possibilités repose sur les mesures de possibilité et de nécessité. La mesure de possibilité d'un événement A parmi l'ensemble des hypothèses considérées Ω est noté Π(A). Elle peut être vue comme une mesure floue, et permet de représenter l'imprécision. Elle a pour propriété :

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \tag{3.5}$$

À cette mesure vient s'ajouter la mesure de nécessité, N(A), qui s'interprète comme la mesure d'impossibilité de l'événement contraire. Elle permet de représenter l'incertitude et a pour propriété :

$$N(A \cap B) = \min(N(A), N(B)). \tag{3.6}$$

Ces deux mesures sont liées par la relation suivante :

$$N(A) = 1 - \Pi(\overline{A}), \quad \forall A \subseteq \Omega. \tag{3.7}$$

Dans de nombreuses applications, on recourt à l'utilisation de méthodes floues pour estimer les degrés de possibilité. La modélisation des incertitudes est alors réalisée de manière implicite, sans que l'on cherche à les modéliser, grâce à la relation (3.5) puisqu'elle permet de déduire une nécessité à partir d'une possibilité.

3.5.4 Fusion par fonctions de croyance

La théorie des fonctions de croyance a l'avantage de modéliser l'information en assignant à chaque observation une masse de croyance m vis-à-vis de plusieurs sous-ensembles A de classes. Soit $\Omega = \{\omega_1, \ldots, \omega_k, \ldots, \omega_K\}$ l'ensemble des K classes considérées, également appelé cadre de discernement. L'incertitude peut être modélisée via cette théorie en assignant à chaque classe, appelée singleton ou hypothèse simple, une masse de croyance de façon similaire à la théorie des probabilités. En outre, l'imprécision peut également être modélisée car la théorie permet d'assigner des masses de croyance vis-à-vis de sous-ensembles $A \subseteq \Omega$ de plus haute cardinalité, pouvant être appelés disjonctions ou hypothèses multiples, parmi le cadre de discernement. L'ensemble des masses affectées aux hypothèses doivent respecter la relation suivante :

$$\sum_{A \subseteq \Omega} m^{\Omega}(A) = 1. \qquad \textbf{(3.8)}$$

Ainsi, la théorie des fonctions de croyance a l'avantage de permettre la modélisation des incertitudes et des imprécisions de manière explicite.

3.6 Méthodes choisies et les raisons du choix

L'application que nous avons mise en œuvre utilise la fusionà base des méthodes non paramétriques, avec une méthodede fusion des décisions basée surle vote de AND (si tous les systèmes ont décidé 1 alors OUI).

La fusion des décisions de plusieurs systèmes permettant d'augmenter la fiabilité globale du système résultant, la méthode à base de vote de AND permettant d'augmenter la précision.

3.7 Fusion de décisions

Ce type de fusion agit au niveau de l'espace de décision. Elle effectue l'association d'informations élaborées (numériques ou symboliques) qui peuvent être considérées comme des propositions de décision. En effet, au lieu de confier la solution d'un problème à un seul algorithme (expert), on tente de diminuer l'erreur globale d'un système en demandant à plusieurs experts simultanément de prendre une décision selon leurs compétences, celles-ci étant ensuite fusionnées pour aboutir à une décision finale.

Nous définissons un système de fusion de décisions comme une boîte noire qui accepte en entrée des sources de décision partielles δ_i; avec $(i \in 1 \ldots . n)$et$\delta_i \in \{0,1\}$, éventuellement munies d'une confiance $\gamma_i$et qui fournit en sortie une décision finale basée sur la combinaison des entrées. La figure (3.1) donne une idée d'un tel système.

Le domaine de la fusion étant extrêmement vaste, nous ne prétendons pas ici le traiter de manière exhaustive, mais nous allons plutôt l'aborder par des cas pratiques qui ont permis d'améliorer les résultats de nos systèmes.

3.8 Fusion par le vote (and)

Le principe du vote est la méthode de fusion d'informations la plus simple à mettre en œuvre. Plus qu'une approche de fusion, le principe du vote est une méthode de combinaison particulièrement adaptée aux décisions de type numériques ou symbolique. Notons $S_j(x) = i$ le fait que la source S_j attribue la classe C_ià l'observation x. Nous supposons ici que les classes C_i sont exclusives. A chaque source nous associons la fonction indicatrice :

$$M_i^j(x) = \begin{cases} 1 & \text{si} \quad S_j(x) = i, \\ 0 & \text{sinon.} \end{cases} \tag{3.9}$$

La combinaison des sources s'écrit par :

$$M_k^E(x) = \sum_{j=1}^{m} M_k^j(x), \tag{3.10}$$

Pour tout k. L'opérateur de combinaison est donc associatif et commutatif. La règle du vote par AND consiste à choisir la décision prise par tous les sources.

3.9 Conclusion

Nous avons étudié les différentes approches de fusion d'informations haut niveau, en faisant ressortir leurs avantages et inconvénients, et notamment la facilité pour chacune d'entre elles à être employées pour des données numériques et symboliques. La plus simple est le vote, mais les approches probabilistes, possibilistes, et par la théorie des fonctions de croyance possèdent un formalisme plus intéressant permettant de tenir compte des imperfections.

Chapitre 4 :Conception

4.1 Introduction

Ce dernier chapitre est consacré à la conception de l'application qui permettra d'identifier des personnes par la reconnaissance faciale et La reconnaissance automatique du locuteur. Le programme sera codé en MATLAB, son rôle sera d'identifier, parmi une base de visages connus et une base de voix connus.
Plusieurs étapes sont nécessaires, l'étape d'extraction des caractéristiques est la plus importante car les performances du système en dépendent (résultats et robustesse, un temps de latence acceptable pour des applications Temps réel).

4.2 Présentation l'outil de développement

4.2.1 MATLAB

MATLAB (MATrixLABoratory) est un logiciel interactif basé sur le calcul matriciel. Il est utilisé dans les calculs scientifiques et les problèmes d'ingénierie parce qu'il permet de résoudre des problèmes numériques complexes en moins de temps requis par les langages de programmation courant, et ce grâce à une multitude de fonctions intégrées et à plusieurs programmes outils testés et regroupés selon usage dans des dossiers appelés boites à outils ou "toolbox".
Son objectif, par rapport aux autres langages, est de simplifier au maximum la transcription en langage informatique d'un problème mathématique, en utilisant une écriture la plus proche possible du langage naturel scientifique [42].

4.2.2 Pourquoi choisir matlab ?

•Il n'y a pas vraiment de langages de programmation standards dans le domaine scientifique, mais MATLAB s'en rapproche.

•Un bon langage interprété pour le travail numérique et la visualisation.

Matlab utilise des fonctions précompilées et un langage basé sur les vecteurs pour faire un travail numérique et de visualisation sans avoir besoin d'être un langage compilé.

• Les opérations vectorielles. L'ajout de deux tableaux en même temps ne nécessite que d'une seule commande, au lieu d'une boucle « for » ou « while ».

• La sortie graphique est optimisée pour l'interaction. Il est possible de tracer des données très facilement, et ensuite changer les couleurs, tailles, échelles, etc., en utilisant les outils graphiques interactifs.

4.3 Implémentation du système de reconnaissance faciale

4.3.1 Présentation des fonctions principales de notre système

La présentation d'une fonction passe par la mise en exergue de son nom et du critère permettant d'évaluer si oui ou non cette fonction a été remplie lors du fonctionnement du système.

Fonction N°1

Nom : Identifier un individu en temps réel.

Critère : Fournir des informations sur un individu détecté dans une image et se trouvant dans la base de données.

Fonction N°2

Nom : Authentifier un individu sur une photo.

Critère : Dire avec un niveau de certitude qu'un individu est bien celui qu'il prétend être.

Fonction N°3

Nom : Apprendre un individu.

Critère : Un utilisateur devra être capable d'ajouter un individu dans la base de données du système à tout moment.

4.3.2 Contraintes d'exécution des différentes fonctions

Contrainte N°1

Nom : Avoir un bon taux de bonne reconnaissance.

Critère : Taux de bonne reconnaissance supérieure ou égale au seuil de référence qui sera déterminé ultérieurement.

Contrainte N°2

Nom : Posséder des images d'entrées de mêmes caractéristiques.

Critère N°1 : Taille de l'image = 103 * 103 pixels

Critère N°2 : Format de l'image : PNG

Contrainte N°3

Nom : Avoir une base d'apprentissage bien fournie.

Critère : Avoir 10 images pour chaque individu dans la base de données de reconnaissance.

4.3.3 Comment faire le prétraitement des images faciales pour la reconnaissance faciale :

Du fait que les images avec lesquelles nous travaillons dans le système proviennent généralement de différentes sources, elles possèdent généralement des fonds différents, des variations de contrastes, des résolutions différentes et sont de tailles différentes. Il nous a ainsi paru nécessaire de mettre à l'entrée du système un module de prétraitement de ces images dont le rôle principal est la normalisation de l'image à son entrée dans le système.

Il est extrêmement important d'appliquer diverses techniques de prétraitement d'image pour standardiser les images que nous fournissons à un système de reconnaissance faciale. La plupart des algorithmes de reconnaissance de visage sont extrêmement sensibles aux conditions d'éclairage de sorte que s'il a été formé pour reconnaître une personne dans une pièce sombre, il ne sera probablement pas la reconnaître dans une salle lumineuse.

Il y a aussi d'autres problèmes, tels que le visage doit être dans une position très cohérente dans les images (tels que les yeux se trouvant dans les mêmes coordonnées de pixels), de taille uniforme, angle de rotation, émotion (sourire, colère, etc). C'est

pourquoi il est si important d'utiliser un prétraitement d'image avant d'appliquer la reconnaissance faciale.
Pour la normalisation de l'image, le module applique les traitements suivants :

a-Transformation en une image a niveau de gris

Cette transformation permet, dans le cas et rien que dans le cas où l'image source est en couleur, de la transformer en une image Noir/Blanc afin de réduire la taille en mémoire de l'image. Car celle-ci est désormais représentée uniquement sur une seule couche et la valeur d'un pixel est comprise entre 0 et 255 car codée sur 8 bits.

b-Étirement d'histogramme

Étirement d'histogramme ou encore égalisation d'histogramme est un traitement d'image permettant de répartir au mieux les intensités lumineuses sur l'ensemble de la plage de valeurs possibles (0 et 255). Ce traitement permet ainsi d'ajuster le contraste de l'image.

c-Application du filtre médian

L'application du filtre médian nous permet d'éliminer un type particulier de bruit, dit «Salt and Pepper noise» qui consiste en des tâches dispersées d'intensité très forte ou très faible. L'utilisation de ce filtre est très importante car malgré la réduction du bruit dans l'image, elle conserve également les contours ; ce qui est très utile pour la détection d'un visage et sa reconnaissance.

d-Redimensionnement de l'image

Il s'agit en fait ici de respectivement réduire ou augmenter la taille de l'image pour qu'elle soit de 103 * 103 pixels.

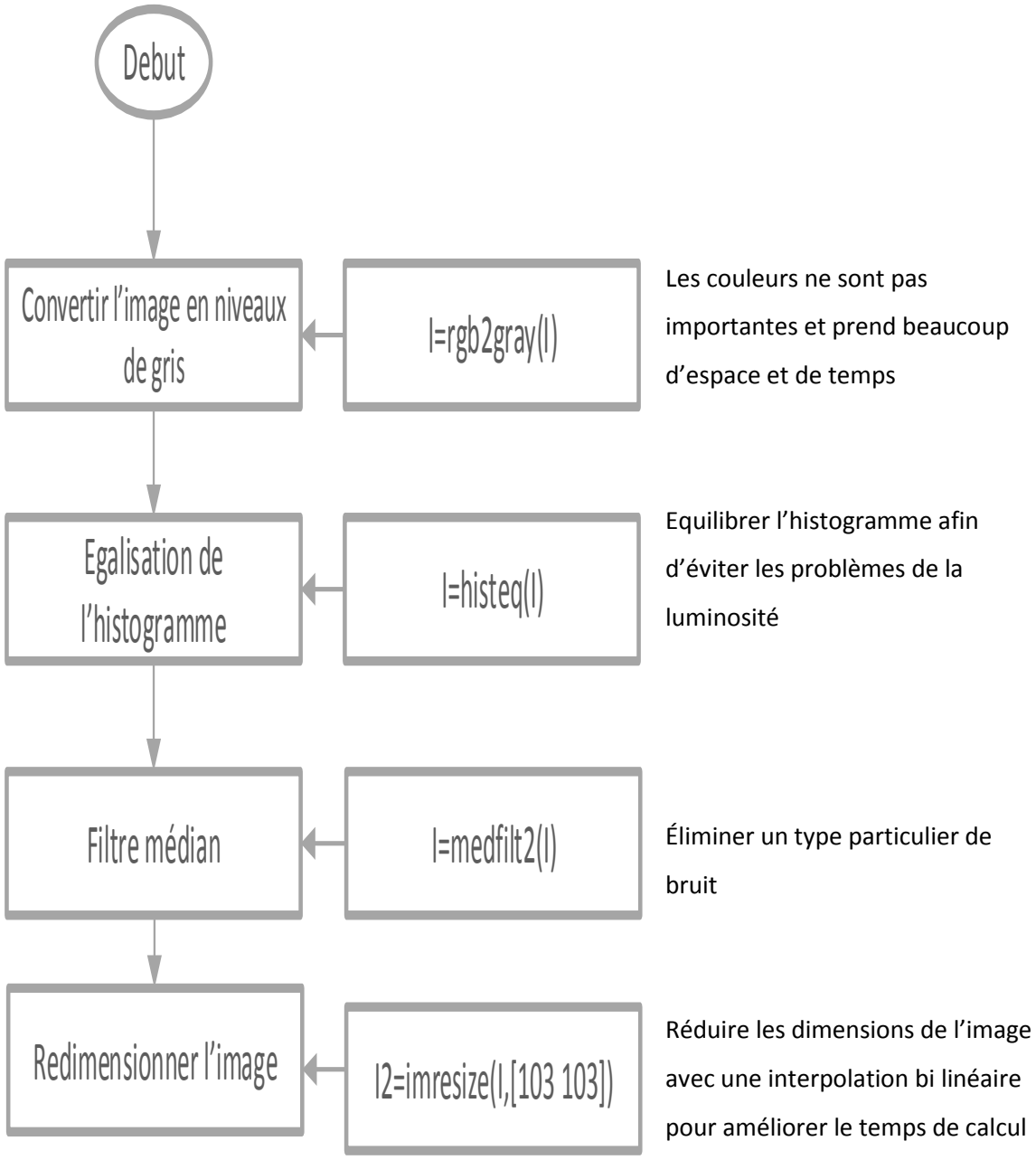

Figure 4. 1 .Organigramme du prétraitement

4.4 Bases de données

Les bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance faciale. Toutefois, ces bases de données sont généralement adaptées aux besoins de quelques algorithmes spécifiques de reconnaissance, chacune d'elle a été construite avec des conditions

d’acquisitiond’images de visages diverse (changements d’illumination, de pose, d’expressions faciales) ainsi que le nombre de sessions pour chaque individu.

Mentalement pour créer une base de données à partir d’images d’apprentissages, on va créer le répertoire pour enregistrer les fichiers d’images et quelle personne de chaque fichier image représente. Par exemple on pourra mettre cela dans un répertoire : **['C:\Users\client\Documents\MATLAB\ApplicationFinal\database\' a1]**

a1 : nom de personne de chaque fichier image.

Pour créer la base de données à partir de ces images, nous utilisons la fonction «**fullfile**" puis "**imwrite**".

Le programme peut alors charger tous dans un ensemble d’images en utilisant la fonction "**imageSet('database','recursive')** ".

Figure 4. 2 . Base de données

C’est une base d’images non bruitées enregistrées avec des prétraitements. Chaque individu est représenté par un ensemble de 10 images regroupant les différentes

expressions faciales, ces images sont au niveau de gris et ont une taille de 103X103 pixels.

4.5 Fonctionnement du bloc «detection du visage»

Le bloc détection du visage, comme son nom l'indique a pour principal objectif de détecter un ou plusieurs visage(s) sur une image. Et, pour le faire, il utilise La méthode de Viola et Jones.

Pour effectuer cette cruciale tâche, de détection de visage, nous avons utilisé le classificateur "**vision.CascadeObjectDetector**" disponible dans MATLAB. Toutefois, elle ne détecte dans la majeure partie des cas que les visages frontaux, c'est pour cette raison, que l'angle de prise de vue de l'image doit être le meilleur possible (prise de vue frontale).

La figure ci-après montre un exemple de détection de visage sur une image.

Figure 4. 3 . Exemple de détection de visage.

4.6 Fonctionnement du bloc « reconnaissance du visage »

Maintenant que nous avons une image faciale prétraitée, nous pouvons effectuer la méthode basée sur des caractéristiques d'apparence locales pour la reconnaissance faciale. MATLAB est livrée avec la fonction "**extractHOGFeatures**", qui effectue l'opération de l'extraction des caractéristiques locales de HOG, mais nous avons besoin d'une base de données (ensemble d'apprentissage) des images.

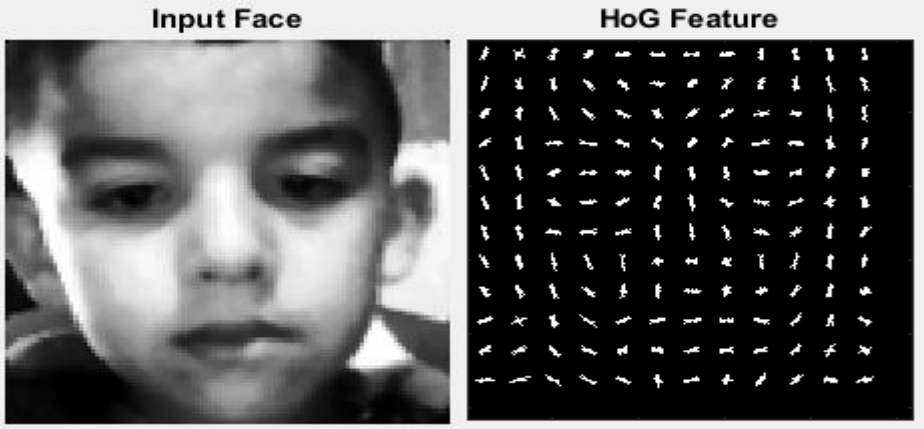

Figure 4. 4 .Exemple de l'extraction des caractéristiques locales de HOG

Le visage est détecté est subdivisée en blocs de tailles égales et ces derniers sont également subdivisés à leur tour en cellules. Pour chacune des cellules, une analyse des gradients des pixels est accomplie afin de former un histogramme de gradient à neuf bandes.

La fonction "**Fitcecoc**" ajusté les modèles multi classes pour les machines vectorielles de support (SVM) pour classer les vecteurs descripteurs obtenus de l'étape d'extraction de caractéristiques.

Donc, nous devons rassembler un groupe d'images faciales prétraitées de chaque personne que nous voulons reconnaître. Par exemple, si vous voulez reconnaître quelqu'un d'une classe de 10 étudiants, alors vous pouvez stocker 10 photos de chaque personne, pour un total de 100 images faciale prétraitées de la même taille. De plus, en créant un classificateur unique pour chaque nouvelle personne, il devient possible de rendre ceux-ci beaucoup plus spécialisés à la reconnaissance d'une personne donnée, ce qui est très important dans le contexte de la reconnaissance de visages. En revanche, en utilisant "**predict**" pour valider la reconnaissance.

La décision globale de reconnaissance d'un individu spécifique dans le système peut être accomplie suite à la classification en considérant "**corr2**"le coefficient de corrélation r entre l'image du visage de l'individu proclamé et l'image du visage de l'individu à authentifier en fonction d'un seuil de reconnaissance appliqué sur chacun des résultats de classification.

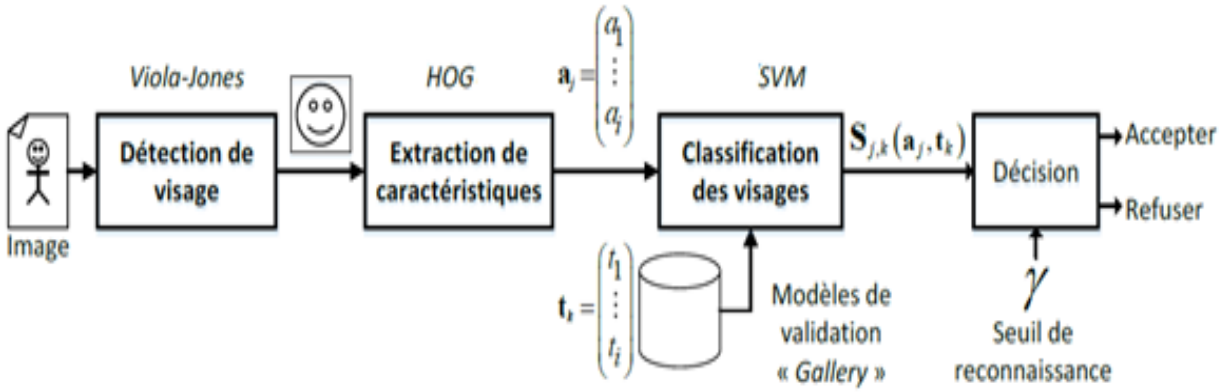

Figure 4. 5 .Schéma-blocs adapté au système de reconnaissance de visages

4.7 La mise en œuvre de reconnaissance faciale en temps réelà partir d'une caméra :

Il est facile d'utiliser un flux de webcam comme entrée pour le système de reconnaissance de visage au lieu d'une liste de fichiers. Fondamentalement, il suffit de saisir les images d'une caméra au lieu de puiser dans un fichier, puis de maintenir l'exécution jusqu'au départ de l'utilisateur, au lieu de simplement exécuter jusqu'à ce que la liste des fichiers est terminée. MATLAB offre la fonction"**webcam ()**".
Maintenant, on a une façon de reconnaître les gens en temps réel à l'aide d'une caméra, mais pour qu'il saisisse de nouveaux visages, il faut arrêter le fonctionnement de la reconnaissance du visage, enregistrez les images de la caméra en tant que fichiers images, mettre à jour la liste des images d'apprentissage, puis exécutez à nouveau le programme en mode reconnaissance du visage en temps réel.

4.8 Implémentation du système de reconnaissance automatique du locuteur

Pour développer notre système d'identification automatique de locuteur, nous allons passer par trois phases : la para métrisation, la modélisation et finalement l'étape de la décision.

4.8.1 Technique proposée d'extraction des paramètres

Nous allons commencer par la méthode MFCC pour l'extraction des paramètres. L'objectif est d'utiliser des paramètres appropriés pour la reconnaissance du locuteur. Le signal acoustique contient différents types d'informations à propos du locuteur. Les MFCC sont utilisés en reconnaissance de la parole et en identification du locuteur car ces paramètres sont bien adaptés au signal parole. Pour cela, dans ce travail, nous avons utilisé les paramètres MFCC (18 Coefficients MFCC).

•Coefficients MFCC :

Pour une séquence $s\ (n)$qui caractérise un signal parole :

$$S = \begin{bmatrix} S_0 & S_1 & \ldots\ldots & S_{N-1} \end{bmatrix}^T \quad \textbf{(4.1)}$$

Où N: représente le nombre d'échantillons de signal parole.
Les coefficients MFCC de signal parole en entier peuvent être représentés par :

$$c = \begin{bmatrix} c_0 & c_1 & \ldots\ldots & c_M \end{bmatrix}^T \quad \textbf{(4.2)}$$

Où : M représente le nombre de trames, tandis que$c_0 c_1 \ldots\ldots c_M$, représentent les MFCCs pour toutes les trames représentant le signal parole$0\ ,1\ldots\ldots M.\ c_0$Représente les MFCCs pour la trame numéro "0" qu'on peut exprimer par :

$$c_0 = \begin{bmatrix} cc_{00} & cc_{01} & \ldots\ldots & cc_{0L} \end{bmatrix}^T \quad \textbf{(4.3)}$$

L : représente le nombre des coefficients MFCC qu'on va adopter (Dans notre travail nous avons considéré $L = 18$). Des équations 4.2 et 4.3, on aura les coefficients MFCC globale pour le signal parole considéré par la matrice suivante :

$$c = \begin{bmatrix} cc_{00} & cc_{10} & cc_{20} & \ldots\ldots\ldots cc_{M0} \\ cc_{01} & cc_{11} & cc_{21} & \ldots\ldots\ldots cc_{M1} \\ & & & \\ cc_{0L} & cc_{1L} & cc_{2L} & \ldots\ldots\ldots cc_{ML} \end{bmatrix} \quad \textbf{(4.4)}$$

4.8.2 Modélisation des locuteurs par la quantification vectorielle (QV)

Chaque trame du signal de parole en entrée est représentée par un vecteur acoustique à 18 dimensions. Le vecteur est constitué des 12 premiers coefficients MFCCs plus $C(0)$ comme composante d'énergie de la trame et leurs dérivées premières et secondes.Une quantification vectorielle non-supervisée par segmentation suivant le principe des k-moyennes est utilisée.

Cette quantification est appliquée sur chaque vecteur acoustique $x$avec son plus proche centroïde Cde moyenne et variance(μ, σ^2). La distance $d(x, C)$entre le vecteur xà 18 dimensions et le centroïde Cest basée sur le logarithme d'une fonction de densité de probabilité $D(x, C)$(*Probability Density Function*, PDF) telle que :

$$\ln D(x, C) = -\frac{1}{2}\left[\sum_{j=1}^{18} \frac{(x_j - \mu_j)^2}{\sigma_j^2} + \ln\left((2\pi)^{18} \prod_{j=1}^{18} \sigma_j^2\right)\right] \quad \textbf{(4.5)}$$

Dans le cas présent, les centroïdes sont indépendants les uns des autres. On considère alors que la variance est ici locale au centroïde associé au vecteur considéré. Donc en supprimant les coefficients pondérateurs et les termes constants, nous choisissons d'utiliser la distance simplifiée $d(x, C)$suivante :

$$d(x, C) = \sum_{j=1}^{18} \left(\frac{(x_j - \mu_j)^2}{\sigma_j^2} - \ln \frac{1}{\sigma_j^2} \right) \quad \textbf{(4.6)}$$

Le premier terme de cette équation (IV.6) correspond à une mesure de distance de Mahalanobis entre le vecteur $x$et le centroïde Cde moyenne et variance(μ, σ^2). En effet, on peut exprimer cette distance de Mahalanobis sous la forme d'une distance euclidienne centrée normalisée $d_M(x, C)$:

$$d_M(x, C) = \sqrt{\sum_{j=1}^{18} \frac{(x_j - \mu_j)^2}{\sigma_j^2}} \quad \textbf{(4.7)}$$

Donc compte-tenu de l'expression de cette distance de Mahalanobis$d_M\ (x, C)$, notre distance de PDF simplifiée $d(x, C)$peut alors s'écrire ainsi :

$$d(x, C) = d_M(x, C)^2 - \sum_{j=1}^{18} \ln \frac{1}{\sigma_j^2} \quad \textbf{(4.8)}$$

4.8.3 Phase de décision

C'est la tâche de calculer les distances correspondantes entre le vecteur de test et les modèles de la base de données (modélisé par VQ). La phase de décision à savoir l'identification est effectuée par la distance euclidienne.

La distance euclidienne D entre deux vecteurs X et Y sur MATLAB est :

"D = sum((x-y). ^2). ^0.5". **(4.9)**

4.8.4 Description de la base de données

La base de données utilisée dans l'expérience est une base de données basée sur la voix, nous avons fait l'expérience sur n locuteurs, chaque locuteur possède 1 enregistrement, la durée de ces fichiers est 2 secondes avec une fréquence d'échantillonnage de 44.1 KHz, nous avons réservé 1 enregistrement d'une durée d'une 2 secondes pour le test, les fichiers wav sont nommés de la façon suivante : si.wav avec i le numéro de fichier et varie entre 1 à n.

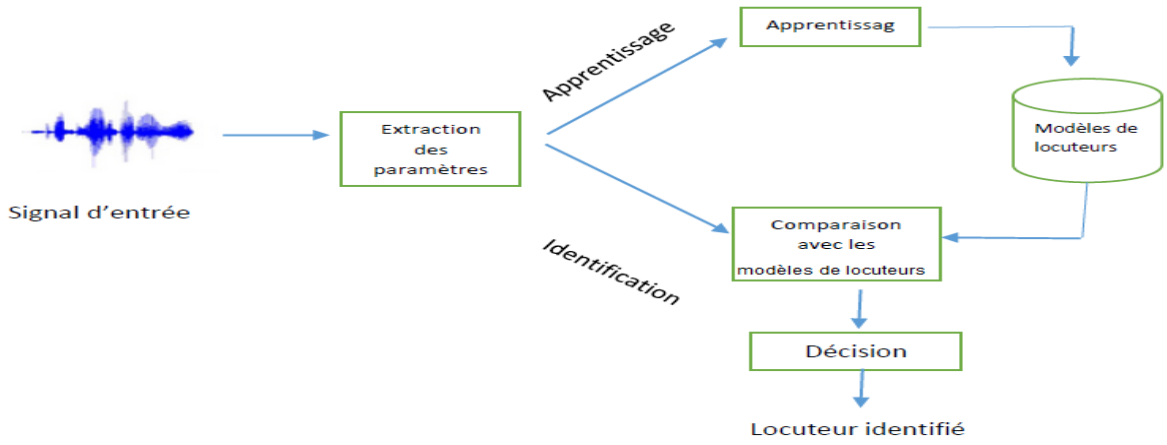

Figure 4. 6 . Processus d'identification

Conclusion

Dans ce chapitre, nous avons présenté les différents systèmes que nous avons utilisés durant la conception et la réalisation de notre projet.

5

Chapitre 5 :Simulation et Résultats

5.1 INTRODUCTION

A travers ce chapitre, nous allons tout d'abord présenter successivement les différents résultats obtenus .Ensuite, les observations que nous avons effectuées et enfin la présentation de l'application que nous avons conçue.

5.2 Configuration des paramètres du système:

5.2.1 Extraction des paramètres faciaux :

Nombre d'image par utilisateur	10
Nombre de coefficients HOG	36
Dimension d'image	103x103
Dimension bloc HOG	8x8

Tableau 5. 1 .Configuration des paramètres faciaux

Pour tester l'efficacité de notre système nous avons créé deux nouvelles bases d'apprentissage :

a-La première base est composée de 10 images d'apprentissages et cela a conduit à des résultats qui sont illustrés dans le tableau suivant :

Test	Taux de reconnaissance (%)	Taux d'erreur (%)
Disponible dans la BD	95	5
Pas disponible dans la BD	91	9

Tableau 5. 2 . *Résultat obtenu sur la 1ere base de données*.

b-La deuxième base est composée de 40 images d'apprentissages, les résultats sont illustrés dans le tableau suivant :

Test	Taux de reconnaissance (%)	Taux d'erreur (%)
Disponible dans la BD	95	5
Pas disponible dans la BD	91	9

Tableau 5. 3 .Résultat obtenu sur la 2emebase de données.

5.2.2 Discussion des résultats :

La méthode HOG que nous avons appliqué donne de bons résultats quand le nombre d'images d'apprentissages est limité.

5.3 Extraction des paramètres vocaux :

Longueur FFT	256
Nombre de coefficient FMCC	18
Nombre de filtres Mel	20
Fenêtrage	HammingWindow

Tableau 5. 4 . Configuration des paramètres vocaux

5.3.1 Estimation du seuil de reconnaissance :

a-Test dans un environnement calme

Disponible dans la BD	Distmin1	Distmin2	Distmin3	Distmin4	Distmin5	Moyenne
HadjAttou	3.7024	4.6100	4.8204	5.5626	4.3865	4.61638
bouizzoul	4.4912	3.8627	3.9401	3.4914	3.8764	3.93236
Fouad	4.0886	4.0493	4.0195	3.6737	4.5718	4.08058

Tableau 5. 5.Estimation du seuil dans un environnement calme.

b-Test dans un environnement bruité

Disponible dans la BD	Distmin1	Distmin2	Distmin3	Distmin4	Distmin15	Moyenne
HadjAttou	8.3664	5.4743	7.5436	6.2305	6.8474	6.89244
bouizzoul	7.5367	6.5560	5.9646	6.3497	7.7126	6.82392
Fouad	7.5737	6.3841	5.9549	6.0888	7.2351	6.64732

Tableau 5. 6 .Estimation du seuil dans un environnement bruité.

c-Test avec des personnes inconnues

Pas disponible dans la BD	Distmin1	Distmin2	Distmin3	Distmin4	Distmin5	Moyenne
HadjAttou	4.1445	4.5575	3.7291	4.2611	4.5540	4.24924
bouizzoul	4.9543	4.4831	4.3496	4.4244	4.3918	4.52064
Fouad	5.6790	4.7178	6.3834	6.1241	4.9431	5.56948

Tableau 5. 7 . Estimation du seuil avec des personnes inconnues.

5.3.2 -Discussion des résultats :

L'identification consiste à reconnaître un locuteur appartenant à une population de plusieurs locuteurs ; on compare pour cela son expression vocale à des références connues ; dans ce but on utilise la distance euclidienne entre son expression vocale et sa référence personnelle. La méthode que nous avons appliquée donne de bons résultats quand l'identification. Mais il donne des distances convergées si l'expression vocale avec sa référence personnelle ou autre, parce que la base de données est modélisée par la quantification vectorielle. Il est donc impossible de fixer un seuil pour la vérification.

5.4 Présentation de l'application

Dans cette partie de ce chapitre, nous allons présenter l'application que nous avons conçue. La présentation d'une application se fera de la manière suivante:
Nous ferons ressortir les caractéristiques matérielles de l'équipement qui pourra l'exécuter :
Présenterons les différentes captures de l'application en cours de fonctionnement et éventuellement son installation :

5.4.1 Environnement du travail

Notre projet a été développé sur un micro portable :

•Processeur : Intel(R) Core(TM) i3-3120MCPU @2.50 GHZ.

•Mémoire Installée (RAM) : 4.00 GO (3.82 GO utilisable).

•Système d'exploitation du PC : Windows 7 Édition Intégrale.

5.4.2 Captures d'écran de l'application en cours d'installation

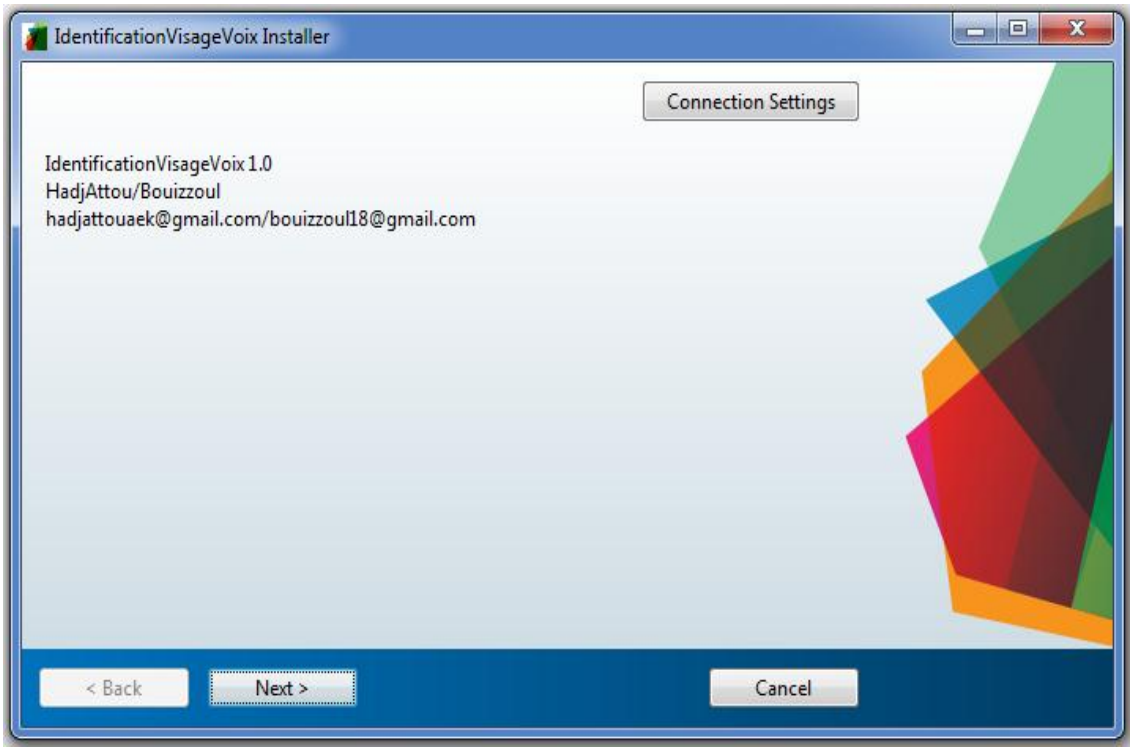

*Figure 5. 1.*Début de l'installation.

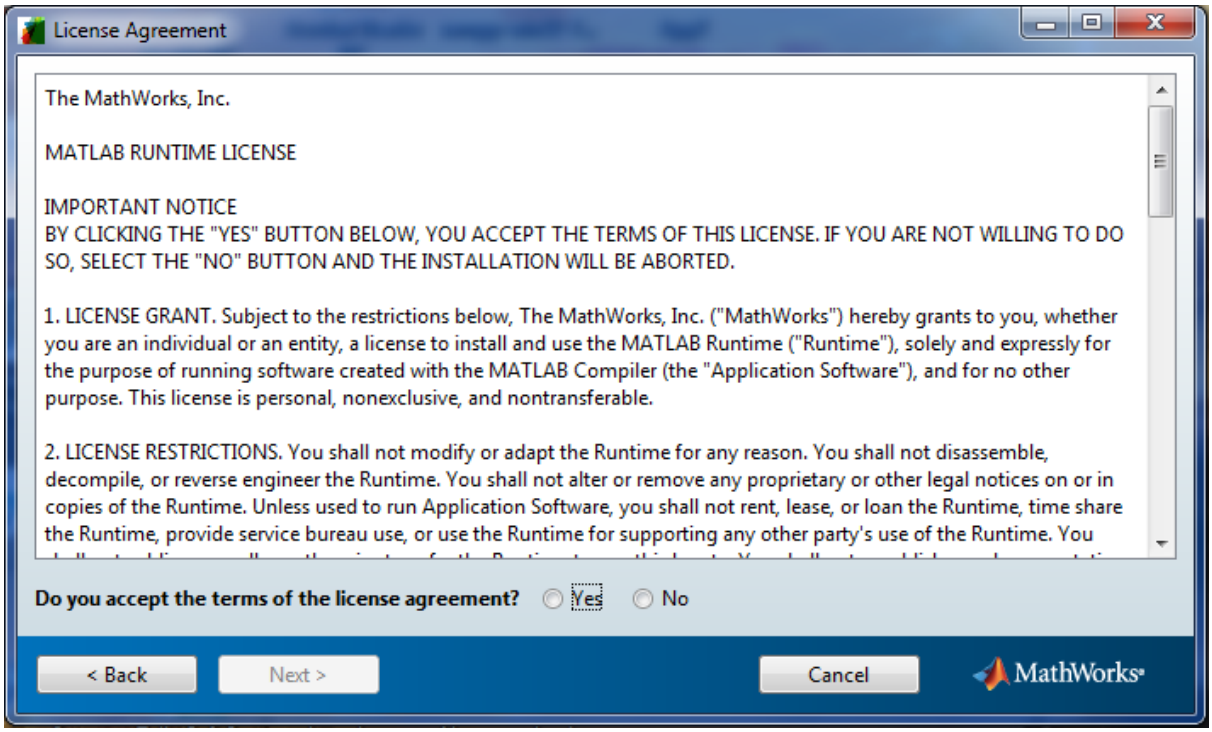

*Figure 5. 2.*Affichage des informations d'installation.

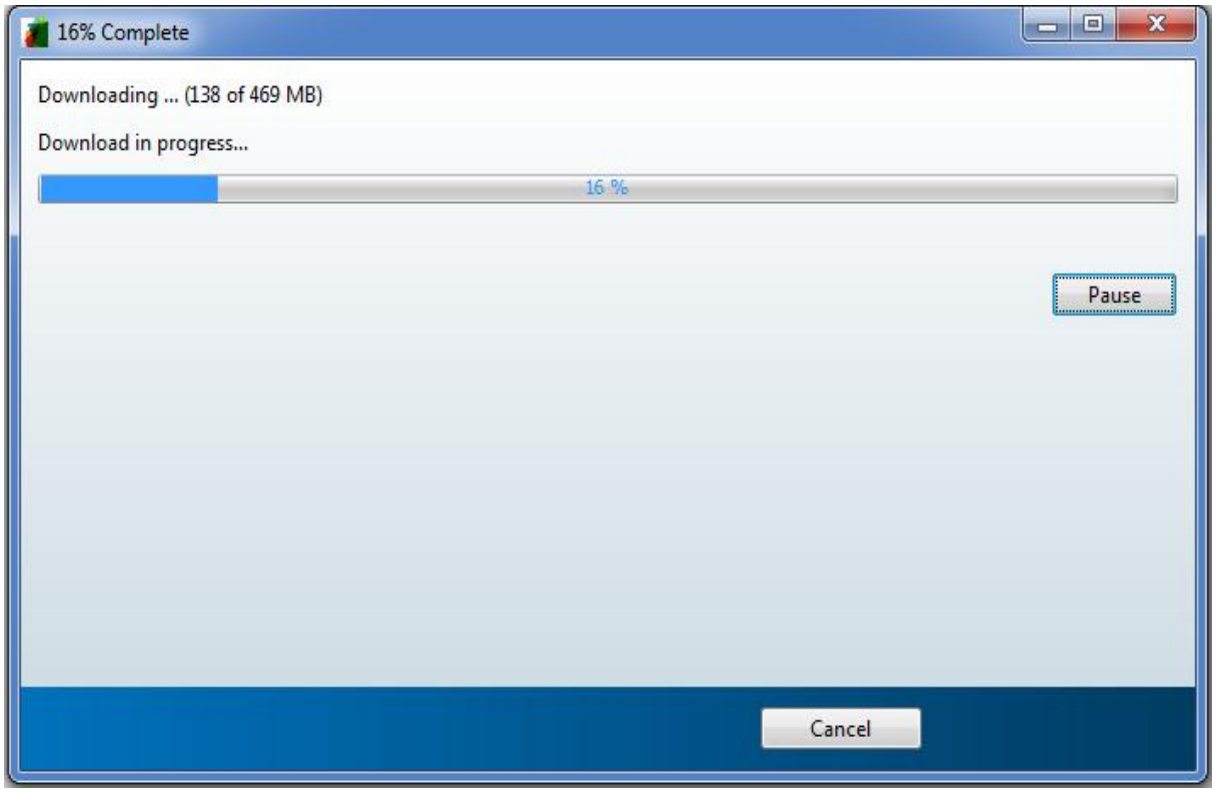

*Figure 5. 3.*Installation de l'application.

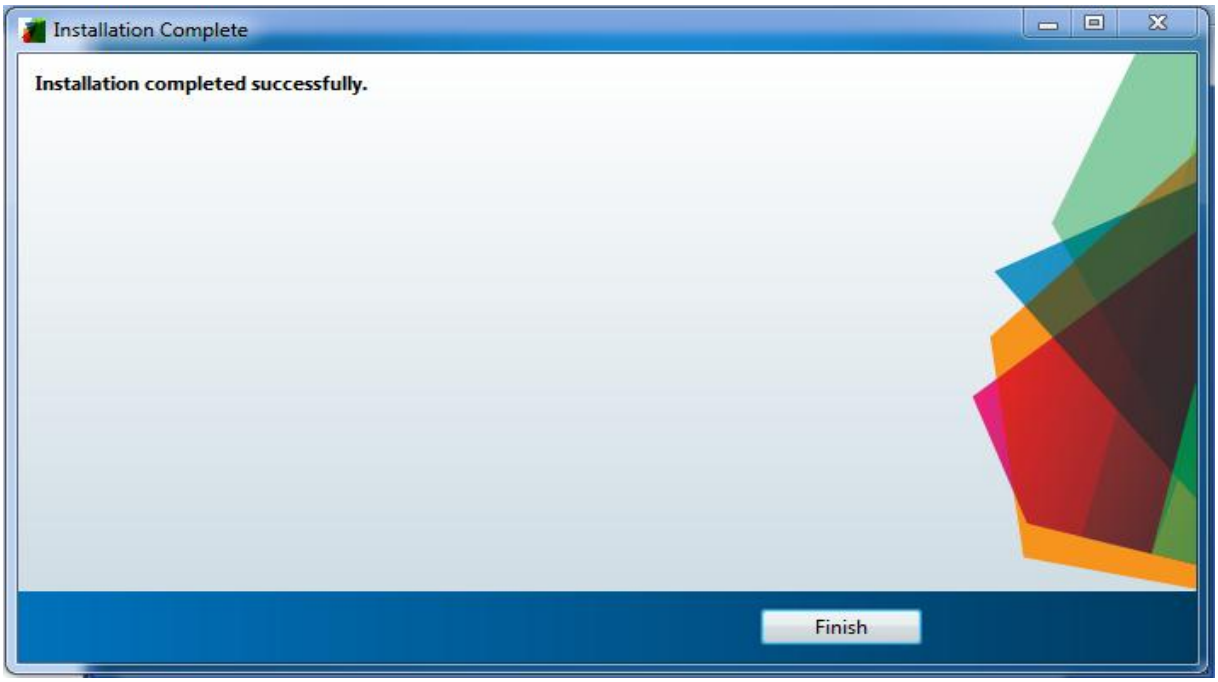

*Figure 5. 4.*Fin de l'installation.

5.4.3 Captures d'écran de l'application en cours de fonctionnement

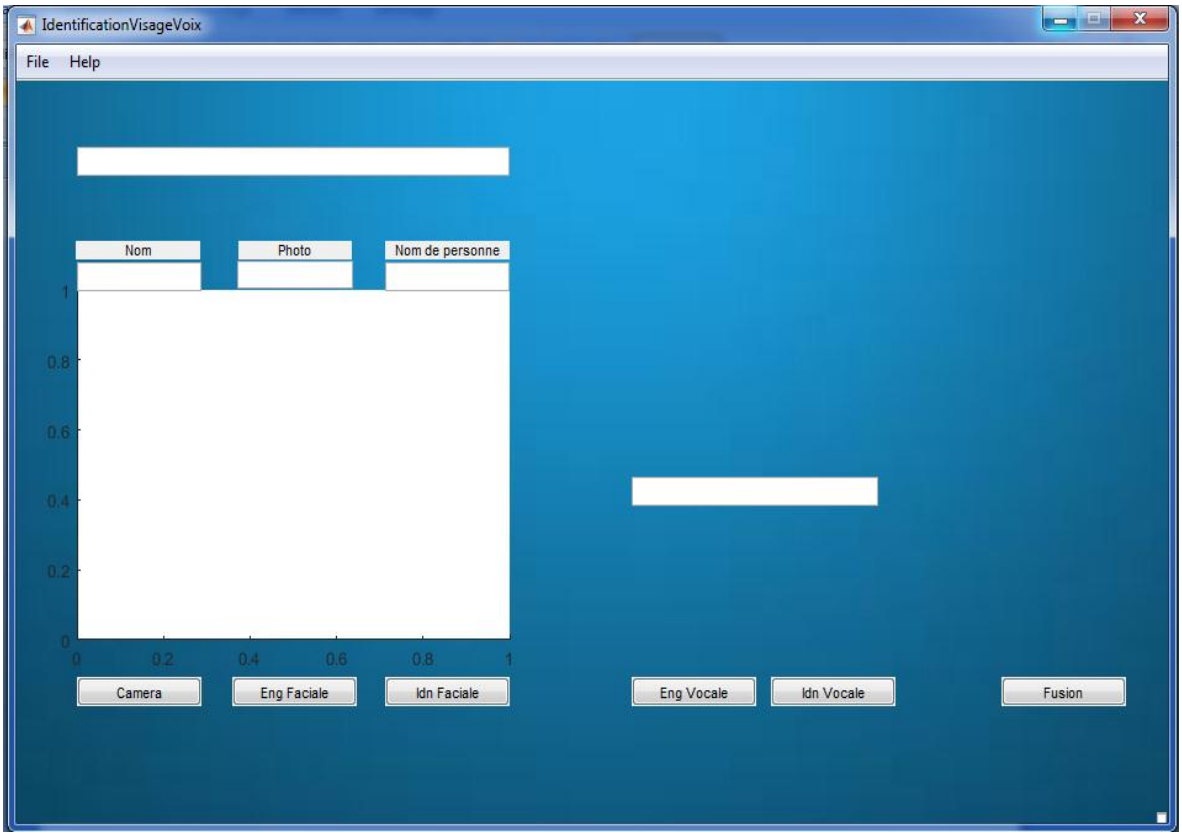

*Figure 5. 5.*Accueil de l'application.

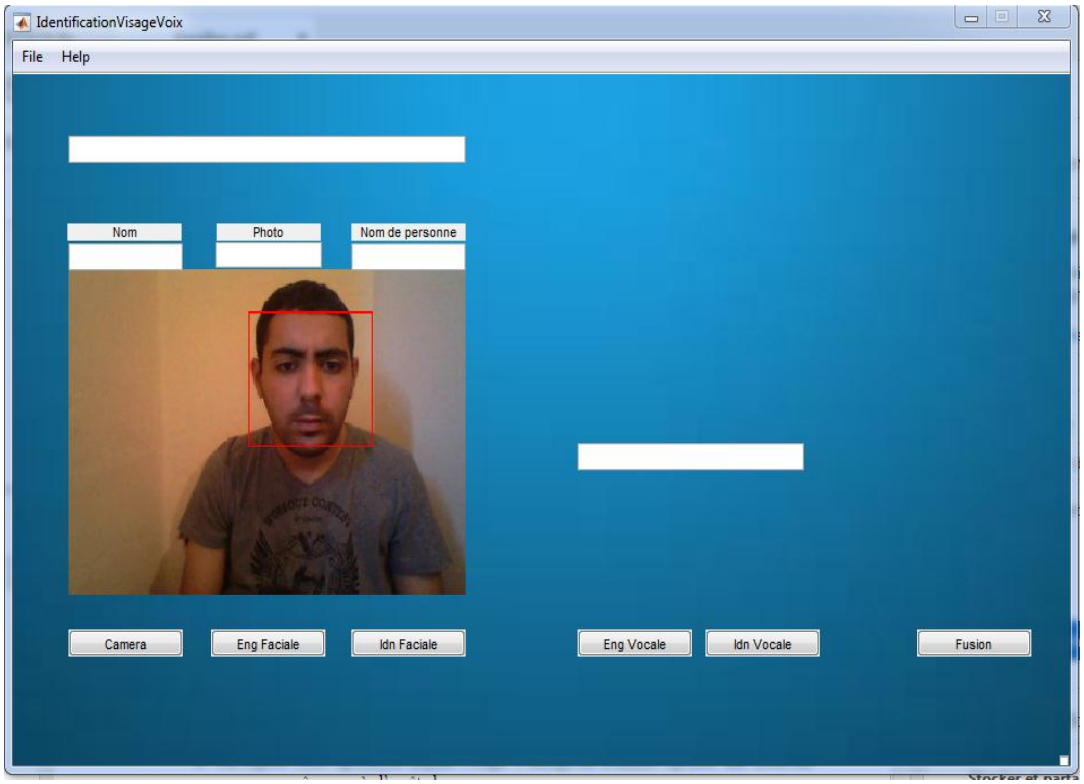

*Figure 5. 6.*La détection de visage automatique.

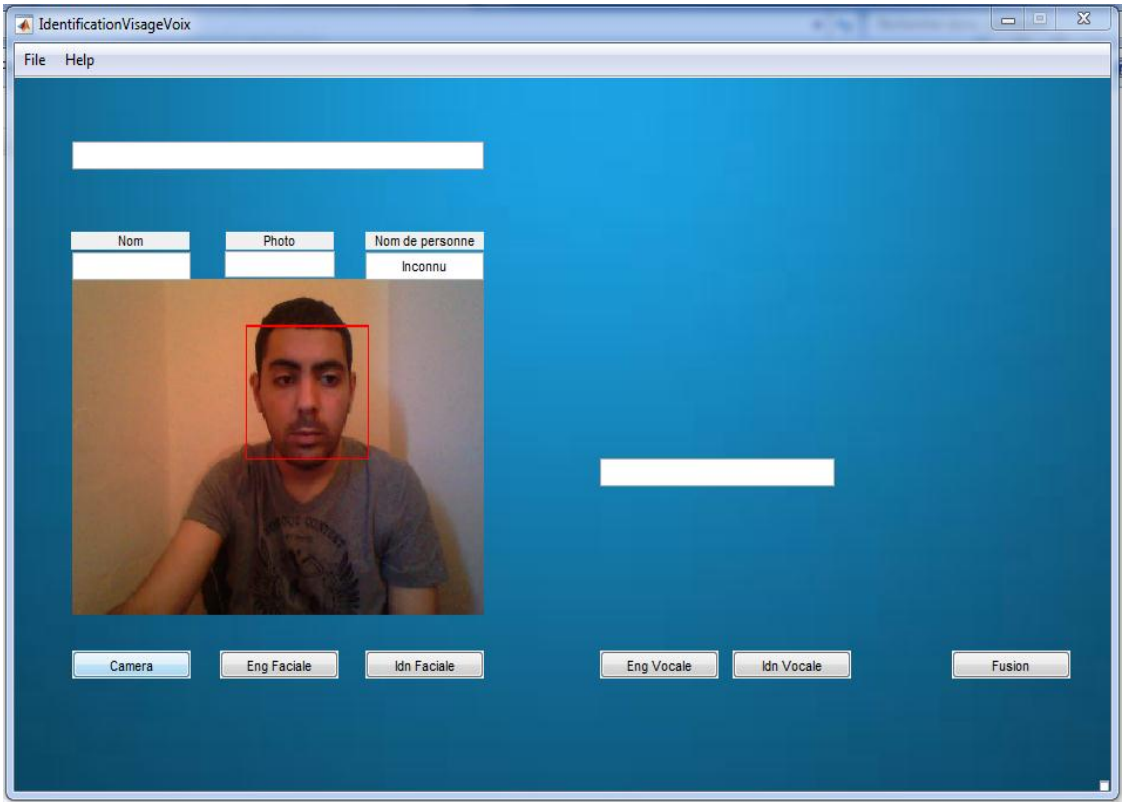

*Figure 5. 7.*Résultat de l'identification en temps réel.

*Figure 5. 8.*Ajouter 10 images à la base de données.

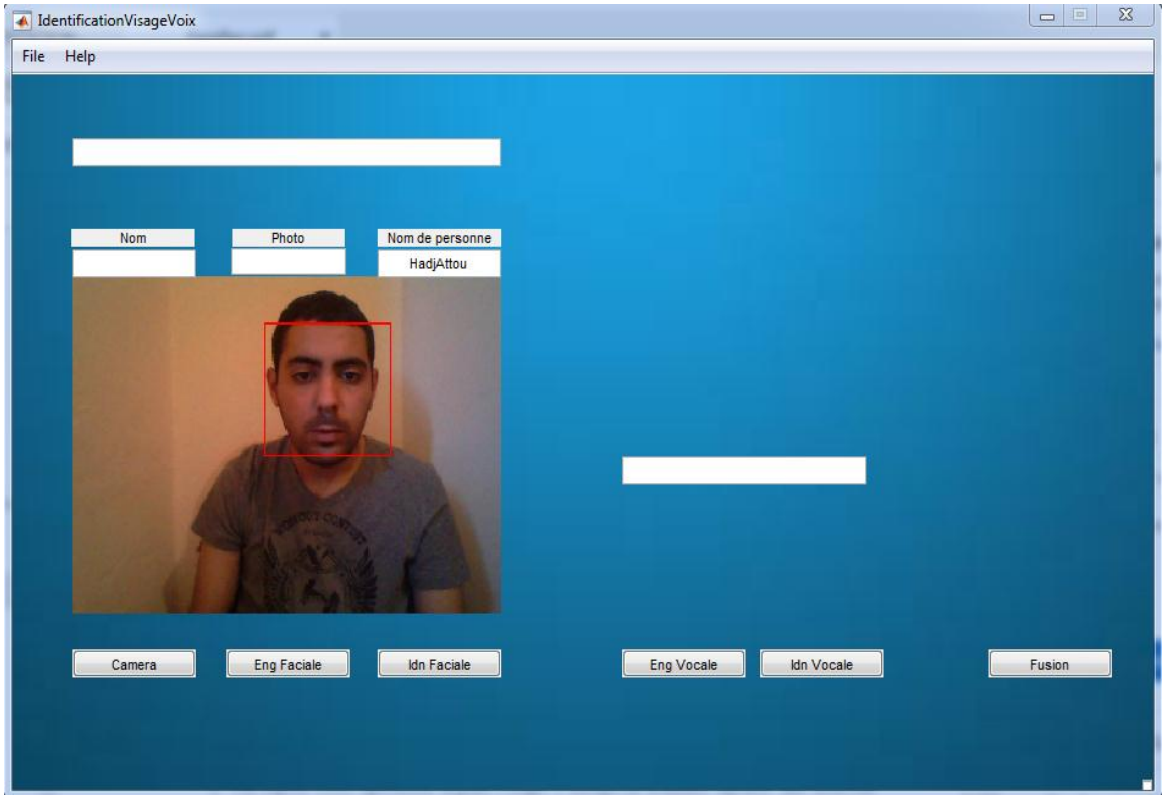

*Figure 5. 9.*Résultat de l'identification en temps réel.

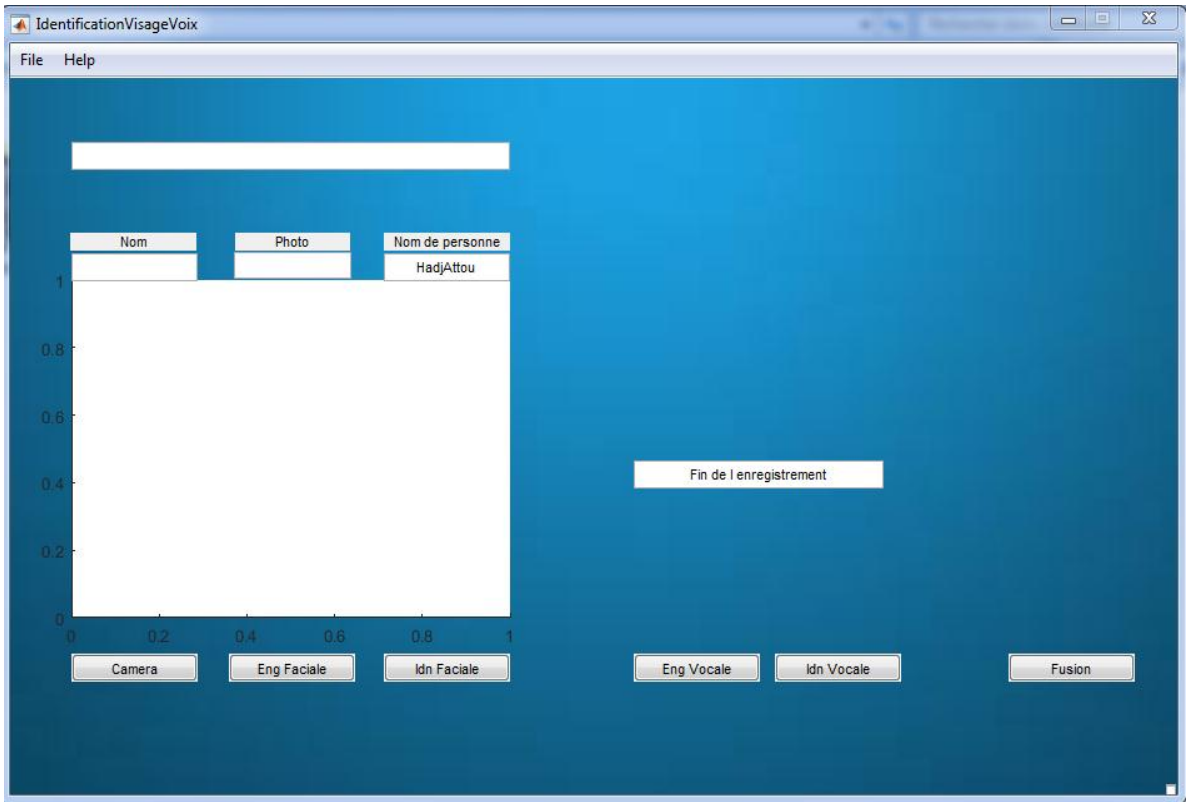

*Figure 5. 10.*Résultat de l'identification vocale.

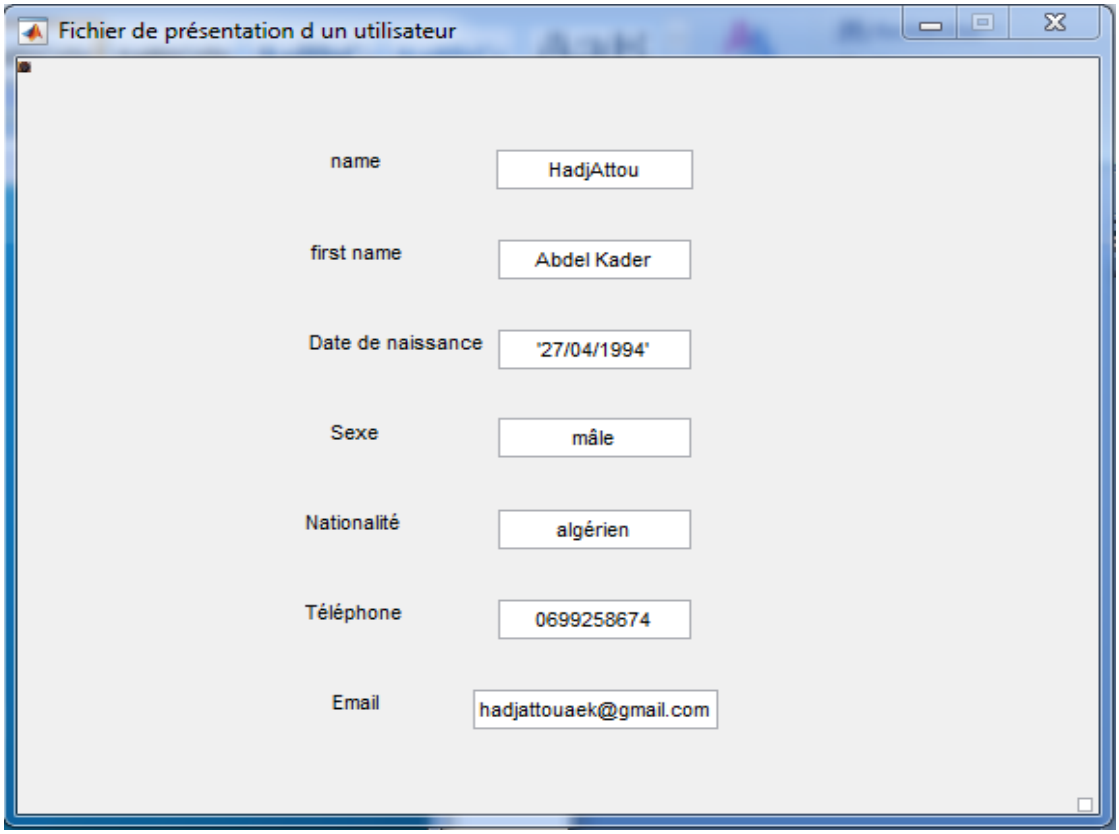

*Figure 5. 11.*Résultat de la fusion.

5.5 CONCLUSION

Dans ce chapitre, nous avons illustré l'architecture globale de notre système d’identification et les détails des fonctions qui le composent, ainsi que le langage qui assure son fonctionnement.

Enfin, Il n'est pas nécessaire de définir un seuil pour chaque partie de l’application (la reconnaissance faciale et la reconnaissance du locuteur), puisque la fusion donne plus de précision au système d'identification.

Conclusion générale et perspectives

Ce travail s'inscrit dans le domaine de la reconnaissance automatique d'individus par le biais de leurs visages et voix. Cette reconnaissance consiste à identifier l'identité d'une personne à partir de son visage et sa voix conjointement. Utilisées principalement pour des raisons de sécurité et de confidentialité, Ces systèmes de reconnaissances sont souvent développés dans des applications de télésurveillance, télé service et d'accès à des endroits sécurisés.

Du point de vue applicatif, les systèmes de reconnaissances bimodaux basé sur le visage et la voix ont atteints un stade leur permettent d'être intégrés dans des applications commerciales de grand publique. En outre, la multi modalité a significativement franchit les limites liées aux performances imposées par les systèmes monomodaux.

Notre projet de fin d'étude s'inscrit dans ce contexte. En effet, nous avons conçue et réalisé une plateforme de fusion biométrique avec la méthode de vote basé sur la reconnaissance de visages et de voix ce qui nous a permet de constater la puissance, la robustesse et la simplicité de mise en œuvre. En effet, le système de reconnaissance automatique d'individus capable, en temps réel, de reconnaître les visages et les voix.

En ce sens, la première partie de l'application qui consiste à localiser les visages. Emploie l'algorithme de Viola et Jones, largement reconnu comme méthode fonctionnant en temps réel et fournissant des résultats robustes et fiables.

Ainsi, l'application s'occupe de la reconnaissance des visages localisés avec un algorithme efficace destiné à reconnaître un individu par son visage en utilisant la méthode HOGFeatures qui se base sur l'extraction des caractéristiques locales de HOG.

Dans La seconde partie de l'application, nous avons utilisé l'outil de quantification vectorielle (VQ) au problème de reconnaissance du locuteur. Les meilleurs paramètres

qui peuvent caractériser une voix parmi d'autres sont les coefficients MFCC. La méthode de QV s'est avérée robuste en termes de reconnaissance.

Nous estimons avoir réalisé un système répondant à l'objectif que nous nous sommes fixés au départ, à savoir la mise en œuvre d'un système permettant la reconnaissance d'individus.

Comme perspectives, et dans le but d'utiliser d'autres modèles pour améliorer les performances, les axes suivant pourraient s'avérer intéressants :

•Amélioration des performances des systèmes actuels par une fusion plus évoluée d'informations.

•Exploitation de nouvelles méthodes d'extraction de paramètres et modélisation.

Annexe Algorithme QV (Quantification Vectorielle)

Algorithme de Lloyd-Max

On se place en distance euclidienne et (y_n) représente un nuage de points avec $0 \leq n \leq N$.

a) Initialisation (t = 0)

On se donne un dictionnaire D_0 de taille K.

b) Construction de la partition

On possède le dictionnaire $D_t = \{D_{it}\}_{i = 1,K}$ après t étapes. On cherche la partition qui minimise l'erreur de quantification associée à D_t composée des classes c_{it} : $\min_{i=1}^{K} d(y_n, \mu_{it})$.

L'erreur de quantification vaut : $D_t = \frac{1}{N} \sum_{n=1}^{N} \left[\min_{i=1}^{K} d(y_n, \mu_{it}) \right]$.

c) Test d'arrêt

Si (par exemple) : $(D_{t-1} - D_t) / D_t < \varepsilon$ alors on s'arrête sinon aller en d).

d) Recalcul des « centres de gravité »

A chaque classe c_{it} de la partition, on associe le « centre de gravité » D_t.
Le dictionnaire est désormais composé des nouveaux D_{it}.
On fait t = t+1 et on va en b).

Algorithme LBG (Linde, Buzo, Gray)

On va construire un dictionnaire de taille K, où $K = 2^p$.

a) Initialisation

Choisir le centre de gravité de l'ensemble d'apprentissage.
Soit D_0 ce vecteur. Le dictionnaire est constitué de D_0. Faire K=0.

b) Eclatement « Splitting »

Tous les éléments d (en nombre 2^k) du dictionnaire sont « éclatés » en 2 vecteurs. Ceci se fait par exemple en transformant chaque d en d+ε et d-ε, où ε est un vecteur donné de norme faible.

c) Convergence

On applique l'algorithme de Lloyd-Max avec les 2^{k+1} éléments ainsi constitués comme dictionnaire de départ. On récupère après convergence un dictionnaire de 2^{k+1} éléments.

d) Arrêt

On fait k = k+1.
Si $k > k_0$ fixé à l'avance, alors arrêt. Sinon aller en b.

Le test d'arrêt peut se faire aussi par rapport à un seuil minimal sur la distorsion des données d'apprentissage par rapport au dictionnaire.

Bibliographie

[1] Ericverzanobres et Morpho :" Reconnaissance faciale"

https://www.morpho.com/fr/reconnaissance-faciale.

[2] Mathieu van wambeke : "reconnaissance et suivi de visages et implémentation en robotique temps-réel", mémoire de master, école polytechnique de louvain2009-2010.

[3] Kalghoumanwar : "Gestion des présences via la technologie de reconnaissance faciale ",https://www.memoireonline.com/01/14/8538/m_Gestion-des-presences-via-la-technologie-de-reconnaissance-faciale-0.html , 2011.

[4] Mr. ghaliahmed :"amélioration de la reconnaissance par le visage", mémoire de master, université des sciences et de la technologie d'oranmohamedboudiaf2015.

[5] Cheng-chinchiang, wen-kai tai, mau-tsuen yang, yi-tinghuang, and chi-

jaunghuang : " à novelmethod for detectinglips, eyes and faces in real time

Real-time imaging", 9(4) :277,287, 2003.

[6] Wenlong zheng and suchendra m. bhandarkar : "facedetection and trackingusing a boosted adaptive particlelter", journal of visual communication and image representation, 20(1):9 - 27, 2009.

[7] Paul Viola et Michael Jones : "rapidobjectdetectionusing a boosted cascade of simple features", ieeecvpr, 2001.

[8] Paul Viola and Michael Jones :"robust real-time face detection. international journal of computer vision", 57 : 137-154, 2004.

[9] M. Kolsch et m. Turk :"analysis of rotationalrobustness of hand detectionwith a viola-jones detector", icpr, vol.3, 2004.

[10] Xiaoyang tan, songcanchen, zhi-hua zhou, and fuyanzhan :" face recognition from a single image per person ,«asurvey. Pattern recognition, 39(2006),1725-1745, 2006.

[11] Patricia rayon villela and juan humbertososSaazuela :"face description with local binary patterns" application to face recognition. In micai 2002 : advances in articial intelligence, volume 2313/2002 of lecture notes in computer sciences, pages 282-291. Springer, heidelberg , 2002.

[12] C. Migniot :" segmentation de personnes dans les images et les videos", these de doctorat, universite de grenoble, français, 2012.

[13] Mme. Benhalloukhadidjaep. Benachenhou :"interface design for human pose estimation", memoire de master, universite des sciences et de la technologie d'oran, 2015.

[14] A. Bettahar et f. Saber :" extraction des caracteristiques pour l'analyse biometrique d'un visage ", memoire de master, universitekasdimerbah, ouargla, 2014.

[15] Saidat djemaa et guezizfatiha :"identification des personnes par l'empreinte de l'articulation des doigts", memoire de master, universitekasdimerbahouargla, 2016.

[16] Nziwouechiadjeuwilfried : "implementation d'un systeme de reconnaissance faciale par la technique des eigenfaces sous java et android", 2016.

[17] Khefifbouchra :"mise au point d'une application de reconnaissance faciale ", memoire de master, universiteaboubakrbelkaid – tlemcen, 2013.

[18] Francis c. Migneault, professeur-superviseur : eric granger, universite du quebec, montreal, canada :"evaluation de methodes de reconnaissance de visages pour l'identification d'individus à partir d'une image de reference" francis.charette.migneault@gmail.com, eric.granger@etsmtl.ca

[19] Gilles gonon et fredericbimbot :"de la reconnaissance automatique du locuteur a la signature vocale", 2017.

https://interstices.info/jcms/c_9758/de-la-reconnaissance-automatique-du-locuteur-a-la-signature-vocale

[20] Houda kadi :"la reconnaissance automatique du locuteur par la voix ip", laboratoire des systemes intelligents & applications, 2014.

[21] Ajgouriadh :"reconnaissance automatique du locuteur a travers les canaux digitaux ", these de doctorat, universitemohamedkhider – biskra, 2016.

[22] Sayoud, halim : " reconnaissance automatique du locuteur approche connexionniste".2003. These de doctorat.

[23] Jousse, vincent : " identification nommee du locuteur: exploitation conjointe du signal sonore et de sa transcription". 2011. These de doctorat. Universite du maine.

[24] Zied sakka, abdennaceurkachouri, ahmedmezghani&mounirsamet :"reconnaissance du locuteur par la technique de laquantification vectorielle", (leti) ecole nationale d'ingenieurs de sfaxb.p.w, 3038 sfax, tunisiesakka_zied@yahoo.fr.

[25] calliope :"la parole et son traitement automatique", edite par j.p.tubachi, masson, 1989.

[26] Petitjean, françois. Description des alignements formes par dtw. 2011.

[27] Linde, yoseph, buzo, andres, et gray, robert m : " an algorithm for vectorquantizer design". Communications, ieee transactions on, 1980, vol. 28, no 1, p. 84-95.

[28] Juang, biing-hwang et rabinerlawrence "fundamentals of speech

Recognition". Signal processingseries. Prentice hall, englewoodcliffs, nj, 1993.

[29] Mohamed senoussaoui :"amelioration de la robustesse des systemes de reconnaissance automatique du locuteur dans l'espace des i-vecteurs", these de doctorat, ecole de technologie superieureuniversite du quebec.

[30] M. Matthieu camus :"identification audio pour la reconnaissance de la parole", these de doctorat del'universite paris descartes sorbonne paris cite, 2011.

[31] Dominique genoud, :" reconnaissance et transformation de locuteurs",these de doctorat, l'ecole polytechnique federale de lausanne (epfl).

[32] Mr sayoudhalim :"reconnaissance automatique du locuteur", these de doctorat, usthb, 2003.

[33] Chattiimane et koulahlamwaad :" la classification des feuilles de vigne a base de descripteur histogramme de gradient oriente ", memoire de master, universitekasdimerbahouargla , 2016.

[34] Boch, i. (2003) :" fusion d'informations en traitement du signal et des images", lavoisier (eds),hermes science publication, 2003.

[35] Isabelle.bloch:"fusion d'informations numeriques : panorama methodologique " isabelle blochecole nationale superieure des telecommunications, dep. Tsi, cnrsumr 5141 ltci paris, france - isabelle.bloch@enst.fr.

[36] Benoit lelandais :"fusion d'informations et segmentation d'images basees sur la theorie des fonctions de croyance : application à l'imagerie medicale tep multi-traceurs ", these de doctorat, universite de rouen, 2013.

[37] l. A. Zadeh : "fuzzy sets as à basis for a theory of possibility. Fuzzy sets and systems", 1(1) : 3–28, 1978.

[38] Arnaud martin :"fusion de classifieurs pour la classification d'images sonar. Revue nationale des technologies de l'information", 2005, e5, pp.259-268. <Hal00286591v1>.

[39] Renaud debon, basselsolaiman, jean-michelcauvin et christian roux, " fusion de l'information : fusion de donnees et de modeles appliques a la segmentation d'images echo-endoscopiques", brest cedex, france.

[40] M. Rouidihoussam : " conception et realisation d'une plateforme de fusion biometrique en score a base des machines a vecteurs de support (svm) ", centre de recherche des technologies avancees(cdta),2008.

[41] Chihebamira et bouhalitnasereddine :"reconnaissance de visages par lda", 2003.

[42] Enmarkhalis :"application assistee par ordinateur pour les malvoyants par acquisition et reconnaissance d'images statiques et dynamiques", universite paris 8, 2013.

[43]: R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior,Guide tobiometrics. new-york: springer-verlag, 2003.

[44]: R. O. Duda, P. E. Hart & D. G. Stork, pattern classification, john wile & sons, usa,2001.