Laboratoire de Télécommunications
et Télédétection

Place du Levant, 2
1348 Louvain-la-Neuve – Belgique

AUTHENTIFICATION MULTIMODALE D'IDENTITÉ

Stéphane Pigeon

Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences Appliquées

Jury composé de Messieurs

Luc Vandendorpe (UCL/TELE) - *Promoteur*
Paul Delogne (UCL/TELE) - *Examinateur*
Marc Acheroy (ERM) - *Examinateur*
Gérard Chollet (ENST) - *Examinateur*
Gilbert Maître (EIV) - *Examinateur*
Michel Verleysen (UCL/DICE) - *Examinateur*
Piotr Sobieski (UCL/TELE) - *Président*

Février 1999

Remerciements

Mes études terminées, un premier emploi dans le secteur privé me fit prendre conscience de la richesse et du confort de travail que pouvait offrir... l'université! Mes premiers remerciements vont donc tout naturellement au Professeur Luc Vandendorpe pour m'avoir offert il y a quatre ans, l'opportunité de travailler au sein du Laboratoire de Télécommunications et de Télédétection. Je le remercie aussi, en tant que promoteur de cette thèse, pour la compétence et la rigueur dont il fit preuve et pour la grande liberté d'action qui me fut offerte.

Les Professeurs Marc Acheroy, Gérard Chollet, Paul Delogne, Gilbert Maître, Piotr Sobieski et Michel Verleysen ont aimablement accepté de faire partie de ce jury. Ils ont montré, à l'occasion de remarques judicieuses, leur intérêt pour ce travail. Je tiens à leur exprimer toute ma reconnaissance.

Je garde un excellent souvenir du temps passé au laboratoire. L'âme d'un laboratoire, c'est avant toute chose, l'ensemble des personnes qui le composent. Qu'elles soient remerciées pour l'aide et l'amitié de nombreuses fois témoignées. Je désire remercier tout particulièrement mes joyeux collègues du A.178, Laurent Schumacher, Benoît Maison, Laurent Cuvelier et François Deryck pour l'excellente ambiance qui régna quotidiennement dans notre bureau.

Il m'est difficile de citer l'ensemble des personnes à qui je dois tant. Je ne peux néanmoins terminer sans exprimer ma profonde gratitude envers mes parents pour leur soutien constant.

Je dédie ce travail à la jeune maman qui a fait de moi un heureux papa!

Stéphane
le 20 janvier 1999

Table des matières

Abréviations

EF	*Expert Frontal*
EP	*Expert Profil*
EQM	*Erreur Quadratique Moyenne*
EV	*Expert Vocal*
FA	*Fausse Acceptation*
FR	*Faux Rejet*
M2VTS	*Multi Modal Verification for Teleservices and Security applications*
MPNG	*Modalité Profil Niveaux de Gris*
MRP	*Modalité Relief du Profil*
PC	*Profil Complet*
PHGH	*Projection Horizontale du Gradient Horizontal*
PP	*Profil Partiel*
PVGV	*Projection Verticale du Gradient Vertical*
TEE	*Taux d'Egale Erreur (TFA=TFR)*
TFA	*Taux de Fausse Acceptation*
TFR	*Taux de Faux Rejet*
$\text{TFR}^{1\%}$	*Taux de Faux Rejet associé à un TFA de 1%*
TS	*Taux de Succès* ($TS = 1 - TFA - TFR$)
TVA	*Taux de Vraie Acceptation*
TVR	*Taux de Vrai Rejet*
VA	*Vraie Acceptation*
VR	*Vrai Rejet*

Notations

ϵ	*Erreur d'expertise* $\epsilon = \|z\|$ *(client) ou* $\epsilon = \|1-z\|$ *(imposteur).*
μ	*Moyenne*
π_A	*Probabilité a priori d'être en présence d'un accès A (Bayes)*
ρ_{ij}	*Corrélation entre experts i et j*
$\sum$	*Matrice de covariance entre experts*
σ	*Ecart-type*
θ_{xy}	*Rotation dans le plan image*
A	*Classe d'accès:* $A = \{c(lient), i(mposteur)\}$
$C_{x\|y}$	*Pénalité encourue de décider x en présence de y (Bayes)*
$CC_i()$	*Courbe caractéristique de l'expert i :* $fr_i = CC_i(fa_i)$
E / $\overline{E}$	*Domaine d'acceptation / domaine de rejet*
$f(FA, FR)$	*Critère d'optimisation du superviseur*
fa_i	*Taux de fausse acceptation relatif à l'expert i (associé au taux de faux rejet* fr_i*)*
FA_*	*Taux de fausse acceptation pour un superviseur de type * (associé au taux de faux rejet* FR_**)*
fr_i	*Taux de faux rejet relatif à l'expert i (associé au taux de fausse acceptation* fa_i*)*
FR_*	*Taux de faux rejet pour un superviseur de type * (associé au taux de fausse acceptation* FA_**)*
k	*Seuil d'acceptation*
(m_x, m_y)	*Centre de masse*
$p^{(i)}_{d\|A}$	*Probabilité que l'expert i prenne la décision d en présence d'un accès A*
(t_x, t_y)	*Vecteur de translation dans le plan image*
$T_Y(z\|A)$	*Fonction de distribution des scores relative à une identité présumée Y pour un type d'accès particulier A*
TFA	*Taux de fausse acceptation obtenu après fusion*
TFR	*Taux de faux rejet obtenu après fusion*
TVA	*Taux de vraie acceptation obtenu après fusion*
TVR	*Taux de vrai rejet obtenu après fusion*
x	*Score (monomodal)*
Y	*Identité présumée*
z	*Facteur d'échelle (Partie 1)*
	Vecteur des scores $[z^{(1)} z^{(2)} ... \ z^{(N)}]'$ *(Partie 2)*
$z^{(i)}$	*Score (normalisé) obtenu au droit de l'expert i*

Introduction

Objet de la thèse

Cette étude traite de l'*authentification* artificielle d'identité par l'analyse du visage essentiellement ainsi que de la voix. Un système d'authentification a pour but de *vérifier l'identité* d'un individu après que celui-ci se soit identifié. Il ne s'agit donc pas d'un système d'*identification* qui lui se charge de découvrir l'identité a priori inconnue d'un individu. Dans ce contexte, nous développerons ou caractériserons divers algorithmes offrant chacun une expertise dans un domaine biométrique particulier: authentification du visage vu de face, authentification du profil et authentification de la voix. Ces algorithmes seront désignés par les termes *experts* et les différents ensembles de caractéristiques biométriques dont il est fait usage par *modalités*. L'information fournie par chaque expert est collectée au *superviseur* qui se charge de fournir une décision finale quant à l'acceptation ou le rejet de la personne que l'on authentifie. La figure 0.1 illustre ces différents concepts.

Une caractéristique essentielle de cette thèse est de reporter la décision d'accepter ou de rejeter un individu au niveau d'intégration supérieur que représente le superviseur. Dans cette optique, les différents experts ne prennent aucune décision et se limitent à transmettre leur avis. Cet avis peut être binaire (acceptation/rejet) ou plus nuancé comme le calcul d'un coefficient de vraisemblance. Dans ce dernier cas, nous verrons comment le fait de maintenir, au niveau des experts, une certaine ambiguïté sur l'identité de l'individu et de reporter la décision ultime d'accès ou de rejet au superviseur, est bénéfique pour le système. Pour obtenir un optimum, une leçon tirée des expériences effectuées dans le domaine de la reconnaissance de formes, consiste à utiliser pour une même interprétation (ici l'identité d'un individu) plusieurs algorithmes élaborés selon des principes différents

(les experts)[60]. Chacun fournit un résultat qui, pris isolément, est parfois peu informatif, mais c'est leur combinaison qui est pertinente. C'est pourquoi, actuellement, de nombreux travaux, dont cette thèse, visent à trouver la meilleure façon de combiner les données produites par différents algorithmes.

Ce travail est divisé en deux parties: la première partie traite de l'étude et de la conception des différents experts tandis que la seconde, de la fusion de données et du superviseur en particulier.

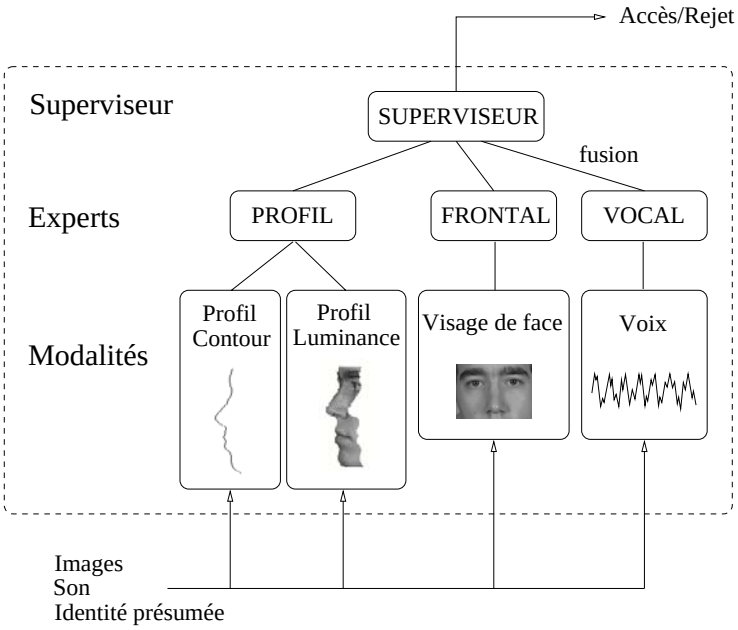


Figure 0.1 - *Le système d'authentification proposé*

Contexte

Avec le développement rapide de nouveaux moyens de communication utilisant les réseaux informatiques comme support de transmission, *Internet* en particulier, de nombreux nouveaux services sont apparus ces dernières années: messagerie électronique, vente à distance, services bancaires à domicile, télé-travail, télé-surveillance, etc. La rapidité avec laquelle ces ré-

seaux et services ont été mis en place n'a pas toujours permis de prendre en considération des aspects importants tels que la sécurité d'accès ou la confidentialité. Cette déficience limite l'essentiel des services offerts aux transactions à confidentialité réduite et généralement non commerciales. Rendre aujourd'hui ces réseaux plus sûrs et pouvoir garantir que nul ne puisse usurper l'identité d'autrui constituent des conditions essentielles à l'élargissement des services offerts. Le but de ce travail est de participer à l'élaboration d'un système capable d'authentifier à distance l'identité d'un utilisateur sur base de ses caractéristiques biométriques. Pour ce faire, des techniques nouvelles, basées essentiellement sur l'analyse du visage mais aussi de la voix, seront utilisées. Au besoin, celles-ci pourront être combinées à des techniques ayant déjà fait leurs preuves, telle l'utilisation d'un mot de passe par exemple, afin d'offrir des performances accrues par rapport à chacune de ces techniques prise séparément. Cette recherche s'inscrit dans le cadre du projet européen M2VTS, décrit au chapitre 3.

De façon générale, les systèmes d'identification ou d'authentification peuvent se classer en deux catégories:

- *les systèmes qui requièrent un contact physique avec l'utilisateur*, tels que la reconnaissance des empreintes digitales, de la rétine ou de la forme de la main. Leur usage est généralement mal accepté, principalement à cause de la nécessité du contact physique (hygiène) et de leur connotation répressive. De plus, de tels systèmes requièrent des capteurs particuliers et ne peuvent être intégrés à coût réduit dans un environnement multimédia classique, faisant essentiellement usage d'un microphone et d'une caméra. Ils offrent néanmoins d'excellentes performances.

- *les systèmes utilisant des capteurs qui ne nécessitent pas de contact direct avec l'utilisateur*, tels que ceux faisant usage de caméras et de microphones. Ces systèmes sont moins coûteux à implémenter (ils peuvent être livrés sous forme logicielle, à destination d'un ordinateur personnel) et sont généralement mieux acceptés par les utilisateurs mais ils ne peuvent prétendre aux mêmes performances que celles offertes par ceux de la première catégorie. C'est pourquoi il est important de pouvoir combiner différentes modalités, comme la parole et l'image par exemple.

Comme il est souhaitable de s'inscrire dans un créneau d'avenir, nous don-

nerons la préférence aux systèmes de vérification d'identité qui peuvent être implémentés sur un ordinateur multimédia et qui font usage de caractéristiques biométriques non contraignantes et naturelles à présenter pour l'individu. *Nous restreindrons donc cette thèse aux algorithmes d'authentification appartenant à la seconde catégorie exclusivement.* Ces derniers sont principalement basés sur l'analyse des images du visage et des échantillons de la voix.

La capacité qu'a l'être humain de pouvoir reconnaître les personnes qui l'entourent, est remarquable tant par la fiabilité des mécanismes mis en jeu, que par la rapidité avec laquelle ils s'exécutent. Ainsi sommes-nous capables de reconnaître des milliers de visages rencontrés durant notre existence et ce, sous diverses conditions d'éclairage, d'environnement, d'expression, de posture ou de vieillissement. Mieux encore, dans certaines situations "hostiles" comme celle d'un garde chargé de la surveillance d'une salle dont l'accès est restreint, nous pouvons distinguer un client d'un imposteur alors que ce dernier fait de son mieux pour ressembler à la personne dont il usurpe l'identité. L'être humain est capable de prouesses insoupçonnées. Offrir de telles performances, similaires en tout ou en partie, tel est le défi lancé en reconnaissance automatique de personnes.

Avant de donner une idée des performances que l'on peut attendre de l'étude de diverses caractéristiques biométriques, signalons combien il est difficile de pouvoir comparer différents algorithmes sur l'unique base de la performance qui leur est attribuée par leur auteurs respectifs ou dans un article récapitulatif comme [13]. Ainsi, alors que certaines bases de données de personnes utilisées pour caractériser ces algorithmes sont représentatives de ce que l'on observe en pratique, d'autres sont plus discutables. Elles fournissent par exemple des images fort semblables les unes des autres, acquises dans un intervalle de temps très réduit voire l'une directement après l'autre. Un algorithme évalué sur de telles bases de données fournit alors des résultats exceptionnels, mais sans signification réelle. D'autre part, des bases de données comme les bases FERET (images statiques) [46] ou M2VTS (séquences d'images et son) [49] permettent de caractériser relativement bien la performance d'un système confronté à des conditions opératoires pratiques. Outre le choix de la base de données, une seconde source de disparité entre les mesures de performances de différents algorithmes, est liée à la façon dont celles-ci sont évaluées, soit en termes d'*identification* ou d'*authentification* comme nous y faisions succinctement référence dans le tout premier paragraphe de cette introduction.

Dans un schéma d'*identification*, aucune information relative à l'identité de la personne qui se présente devant le système, n'est disponible a priori. Les caractéristiques biométriques de la personne à identifier sont comparées à celles de l'ensemble des clients connus du système et stockées dans une base de données centrale. L'identité du client qui offre les caractéristiques les plus proches est alors sélectionnée. Un tel système nécessite un important temps de calcul (de nombreuses comparaisons doivent être effectuées), mais offre une bonne performance en terme de *reconnaissance correcte*: en effet, malgré les variations qui peuvent affecter les caractéristiques physiques d'une personne, celle-ci restera correctement identifiée tant que la distance qui la sépare de son modèle de référence (distance *intra*) est inférieure à celle qui la sépare de tout autre client (distance *inter*). Malheureusement, un tel système ne traite pas le problème d'un éventuel imposteur: celui-ci verra automatiquement son accès autorisé sous l'identité du client qui lui est le plus semblable. Même si l'on fait usage d'un seuil d'acceptation [1], le risque qu'un imposteur soit à même de pouvoir entrer dans le système est d'autant plus élevé que le nombre de clients est grand. Les performances de systèmes d'identification décrits dans la littérature sont souvent proches de 100%. Malheureusement très peu d'information est donnée quant à leur comportement en présence d'accès frauduleux. Brunelli et Poggio [11] rapportent un taux d'identification correcte de 90% en utilisant des caractéristiques géométriques extraites de la vue du visage de face, sur une base de données de 47 personnes. Ce taux augmente jusqu'à 100% pour une méthode de mise en correspondance de fenêtres centrées sur des zones particulières du visage. Moghaddam et Pentland [44] proposent une approche basée sur les visages propres [2] et réalisent 99% d'identification correcte sur les images frontales de 155 individus. Yu et al [68] obtiennent un taux de 100% sur un ensemble de 33 personnes en utilisant diverses caractéristiques extraites du relief du profil.

Dans un schéma d'*authentification*, le candidat qui désire accéder au système doit fournir son identité d'une façon ou d'une autre. Ceci peut être réalisé, par exemple, par l'introduction d'une carte magnétique personnelle ou d'un code en début de processus. Les caractéristiques du candidat sont alors comparées aux caractéristiques de la personne qu'il prétend être. Si celles-ci sont suffisamment proches l'une de l'autre – en d'autres termes si la distance entre ces caractéristiques est en-deçà d'un seuil d'acceptation

1. soit un seuil sur la distance qui sépare les deux visages au delà duquel échoue l'identification.

2. Par référence aux vecteurs propres, en algèbre.

fixé – le client est authentifié. Un tel système bénéficie donc d'un temps de calcul fortement réduit par rapport au schéma d'identification, puisque les caractéristiques biométriques du candidat sont comparées avec les caractéristiques d'un seul client. La performance d'un tel système peut être évaluée en termes de *Taux de Faux Rejet* (TFR), la proportion d'accès clients malencontreusement rejetés par le système, et de *Taux de Fausse Acceptation* (TFA), la proportion d'imposteurs réussissant à y pénétrer. Le *Taux de Succès* (TS) est défini comme étant $TS = 1 - TFA - TFR$ [14]. Goudail et al [27] obtiennent un TS de 93,5% en faisant usage de mesures d'autocorrélations locales sur des images frontales du visage de 116 personnes. Konen et Schulze-Krüger [37] ont développé un système basé sur une extension de la mise en correspondance de grilles élastiques[3], travaillant sur des images du visage de face, et obtiennent un TS de 96% sur 87 individus. Une authentification basée sur le relief du profil développée par Beumier et Acheroy [3] offre quant à elle un TS de 90% sur une base de données comprenant 41 personnes.

Dans le cadre du projet M2VTS, un projet européen traitant de l'authentification multimodale de personnes, différents algorithmes ont été testés sur une base de données commune de 37 personnes [49]. Sur cette base de données, une méthode d'authentification de la parole faisant usage de modèles de Markov cachés[4], offre parmi toutes les modalités prises individuellement, le plus haut TS avec 97,5% [33]. En combinant l'analyse de la parole avec celle du mouvement des lèvres, ce taux augmente jusqu'à 99,4% [33]. Une mise en correspondance de grilles élastiques faisant usage des vues de face issues de cette même base de données offre quant à elle un TS de 89% [19]. En combinant cette dernière méthode avec le module d'authentification de parole mentionné plus haut, le TS augmente alors jusqu'à 99,5% [19].

Ces très bons résultats, atteignant un taux de succès de plus de 99%, sont principalement dus à l'excellente performance de l'authentification de la parole. Selon l'application envisagée, la modalité parole peut ne pas être disponible (reconnaissance de visages sur base de fichiers anthropométriques) ou inutilisable (reconnaissance de personnes dans un environnement bruyant). Cette thèse traitera de ces cas spécifiques et abordera principalement le problème de l'authentification de personnes en utilisant des vues du visage de face et de profil. Nous travaillerons comme suit: à partir de deux modalités liées à la vue du profil, la première étant basée sur le relief du profil et la se-

3. *Elastic Graph Matching*, en anglais.
4. *Hidden Markov Models* (HMM), en anglais.

conde sur la distribution des niveaux de gris le long de celui-ci, nous verrons comment construire un *expert profil* dont les performances seront accrues par rapport à chaque modalité prise séparément. Un deuxième expert fera usage des parties les plus discriminantes du *visage de face*. Différentes techniques de fusion seront alors étudiées et la meilleure approche sera utilisée pour combiner de façon efficace nos deux experts. Ce travail débouchera sur une méthode d'authentification qui, en se basant uniquement sur des images du visage, offrira un TS de 96,5% dans les mêmes conditions de test que celles du projet M2VTS. En y ajoutant l'expert vocal mentionné plus haut, ce taux atteindra 99.95%.

Plan de la thèse

Ce travail est divisé en deux parties. La première introduit et caractérise les performances des différentes modalités et experts dont il est fait usage. La seconde traite de la fusion proprement dite et compare les performances de différents superviseurs tant au niveau théorique qu'en conditions de test pratique. A la suite de ces deux parties nous conclurons cette thèse en résumant les contributions apportées et en envisageant les possibilités de développements futurs.

Plus précisément, la première partie se compose comme suit. Au chapitre premier, nous introduirons deux modalités relatives à la vue de profil. Ces deux modalités donneront par la suite naissance à l'*expert profil*. Dans le deuxième chapitre nous traiterons de l'unique modalité relative à l'image du visage de face. Cette modalité sera par la suite désignée sous le vocable *expert frontal*. Nous présenterons au chapitre 3 les images et les conditions sous lesquelles les performances de ces modalités et experts ont été évaluées. Ces performances seront exposées au chapitre 4. Le chapitre 5 présente un expert n'ayant pas été développé au sein du laboratoire, mais qui complète à merveille l'éventail des caractéristiques biométriques traitées dans le cadre de cette thèse. Il s'agit de l'*expert vocal* basé sur une reconnaissance de la voix. Il sera à son tour évalué sur le même ensemble de test que celui utilisé par les experts profil et frontal. Enfin, le chapitre 6 termine cette première partie par quelques conclusions préliminaires.

La seconde partie quant à elle s'articule autour de cinq chapitres. Les deux premiers chapitres introduisent le problème de la fusion de données et le particularisent au contexte de l'authentification d'identité. Le chapitre pre-

mier formalise le problème d'un point de vue mathématique, tandis que le second propose une approche plus intuitive. Différents superviseurs seront alors étudiés au sein de divers contextes de travail. Le chapitre 3 traite de la *fusion a posteriori*, ce qui revient à comparer différents superviseurs sur base des meilleures performances que l'on peut obtenir après avoir pris connaissance de l'ensemble des données sur lesquelles on effectue le test. Le chapitre 4 propose un scénario plus réaliste où les performances des différents superviseurs sont évaluées sur un ensemble de données inconnu a priori. Dans ce dernier cas, on parlera de *fusion a priori*. Le cinquième et dernier chapitre résume les enseignements apportés tout au long de cette deuxième partie.

Nous arriverons enfin aux conclusions générales. Elles reprennent les diverses contributions apportées tout au long de cette thèse et suggèrent de nouvelles pistes de réflexion utiles dans la perspective de recherches ultérieures.

Première partie

Les Experts

Chapitre 1

La modalité profil

1.1 Contexte

Lorsque l'on désire mettre en œuvre un système d'authentification qui soit à la fois aisé à programmer, rapide d'exécution et qui offre un excellent compromis complexité/performance, la vue de profil représente un maître choix. Contrairement à la vue de face traitée par la suite, la vue de profil, et plus particulièrement le contour du profil, offre des caractéristiques biométriques stables, faciles à isoler et qui ne sont généralement pas accessibles depuis la vue de face. De plus, hormis le cas de la bouche qui peut engendrer des reliefs de profil différents selon que celle-ci est ouverte ou fermée, le contour du profil peut être considéré comme étant relativement indépendant de l'expression du visage.

Les méthodes de reconnaissance du profil peuvent être regroupées en deux familles principales. La première se compose des méthodes basées sur l'extraction de points particuliers du profil, points qui serviront à calculer la similitude qui existe entre le profil candidat et un modèle de référence donné [28, 29, 68]. Leur extraction représente une opération délicate qui doit être faite avec soin si l'on veut bénéficier d'une certaine souplesse par rapport aux images d'entrée. Cette extraction requiert souvent l'usage de règles empiriques et constitue le point faible de ces méthodes. Une fois ces points de référence extraits, la procédure d'authentification devient alors extrêmement rapide et consiste à comparer des rapports de distances. Ces méthodes offrent de bonnes performances pour autant que l'extraction des

caractéristiques biométriques soit assez précise et que le nombre de points extraits soit suffisamment élevé en regard de la taille de la population que l'on désire traiter (typiquement 10 points caractéristiques pour discriminer une trentaine d'individus).

La seconde famille de méthodes d'authentification, traite les profils de façon globale. Plutôt que de comparer un nombre limité de points, c'est la totalité du profil qui cette fois est exploitée. Ces algorithmes sont conceptuellement plus faciles à mettre en œuvre mais requièrent un temps d'exécution généralement plus long que les méthodes précédentes, suite notamment aux normalisations de translation, rotation et facteur d'échelle nécessaires avant de pouvoir comparer deux profils entre eux. Certaines normalisations peuvent néanmoins être évitées en faisant appel à des représentations particulières du contour du profil, telles celles fournies par une analyse de Fourier [1] ou un calcul de courbure locale [3].

Grâce à l'utilisation d'algorithmes rapides, nous avons pu choisir dans le cadre de ce travail, une méthode globale travaillant directement sur les coordonnées x-y du contour du profil [50]. Ainsi, les performances obtenues dépendront uniquement de la capacité qu'offre la vue du profil à distinguer des visages différents et non du choix d'un ensemble particulier de points caractéristiques. A cet avantage s'associe malheureusement un inconvénient: contrairement aux techniques faisant usage de caractéristiques locales, l'utilisation du profil dans sa globalité ne nous permet pas aisément de pondérer différemment les diverses parties du visage, accordant un poids accru aux zones les plus robustes et/ou les plus discriminantes. Une telle technique eut été susceptible d'améliorer les performances liées à la modalité profil.

1.2 Authentification du relief du profil: mise en correspondance du chanfrein

La méthode du chanfrein[1], a pour but de rechercher la meilleure correspondance entre deux images binaires[2]. Un ensemble de transformations géométriques est appliqué à l'une des deux images, appelée *image can-*

1. *Chamfer matching*, en anglais.

2. Une image binaire est une image formée uniquement de points noirs (0) ou blancs (1), sans aucune nuance de gris.

didate, afin de minimiser une distance mesurée par rapport à la seconde, l'*image de référence*. Ces images binaires sont généralement obtenues à partir de contours présents dans les images à traiter, soit le contour du profil dans le cas particulier qui nous intéresse. La mesure de distance utilisée se doit d'être suffisamment précise pour pouvoir aboutir à une bonne mise en correspondance; elle doit également offrir une complexité raisonnable afin de bénéficier d'un temps de calcul réduit. Notre choix s'est porté sur la métrique du chanfrein [51], détaillée au point suivant.

1.2.1 Distance du chanfrein

Le calcul d'une distance euclidienne peut rapidement devenir une opération fastidieuse. Pour une image binaire de taille $k \times l$ et comprenant N points non nuls, le nombre d'opérations nécessaires au calcul de la carte de distance qui lui est associée[3] sera proportionnel à N, qui représente en quelque sorte le contenu de l'image[4]. Ainsi préfère-t-on, en traitement d'images, remplacer le calcul de la distance euclidienne par une approximation qui permet d'éviter cette proportionnalité et qui bénéficie par conséquent d'un temps de calcul réduit. Aussi, il n'est pas toujours nécessaire de recourir à la précision offerte par le calcul de la distance euclidienne. Les caractéristiques présentes dans les images à traiter, servant à fournir les éléments non nuls de l'image binarisée, sont souvent affectées par du bruit (voir section 1.4). Il devient alors vain de vouloir calculer une distance exacte par rapport à des caractéristiques qui ne peuvent être localisées très précisément.

Le terme *chanfrein* faisait initialement référence à un algorithme de calcul de carte de distance basé sur le double parcours de l'image binaire de départ (voir section 1.2.2). Cet algorithme, développé par Rosenfeld et Pflatz [56] et amélioré par la suite par Borgerfors [7], se base sur la propagation d'un ensemble fini de distances locales, définies à l'intérieur de ce qu'on appelle un *masque de voisinage*. Par extension, le terme *chanfrein* désigne aujourd'hui la famille des transformations métriques basées sur une telle propagation de distances locales.

Considérant un masque de voisinage de dimensions 3×3 tel que celui uti-

3. Une carte de distance associée à une image binaire assigne en chacun de ses points la valeur de la plus petite distance qui sépare ce point d'un point non nul de l'image binaire de départ. Voir section 1.2.2.

4. Ce nombre d'opérations est en fait proportionnel à $(kl - N)N$, soit une complexité $O(Nkl)$.

lisé dans le cadre de ce travail et illustré à la figure 1.1, deux distances élémentaires y sont définies: une distance axiale a et une distance diagonale b. En pratique, (a, b) sont choisis entiers afin de bénéficier de la rapidité d'exécution du calcul en valeurs entières. La distance globale obtenue en fin de processus est alors divisée par une constante de normalisation q afin d'approcher au mieux la distance euclidienne.

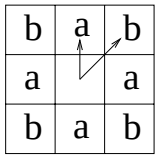


Figure 1.1 - *Distances locales définies dans un masque de voisinage* 3×3

Cette métrique est généralement notée *métrique du chanfrein* $(a, b)/q$. Les valeurs $(3, 4)/3$ sont usuelles et fournissent une approximation satisfaisante de la distance euclidienne tel qu'illustré à la figure 1.2, avec une erreur maximale de 6% [12].

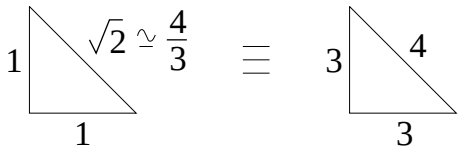


Figure 1.2 - *Choix des paramètres* $(a, b)/q$

Une mesure de la distance globale peut être obtenue en propageant les distances locales a et b suivant respectivement les axes principaux et diagonaux. Cette propagation est illustrée à la figure 1.3.

Ainsi, la distance $d(i, j)$ associée au point de coordonnées (i, j) est directement liée aux distances voisines par la relation:

$$d(i,j) = min \left\{ \begin{array}{lll} d(i-1,j-1)+4, & d(i-1,j)+3, & d(i-1,j+1)+4, \\ d(i,j-1)+3, & & d(i,j+1)+3, \\ d(i+1,j-1)+4, & d(i+1,j)+3, & d(i+1,j+1)+4. \end{array} \right\} \tag{1.1}$$

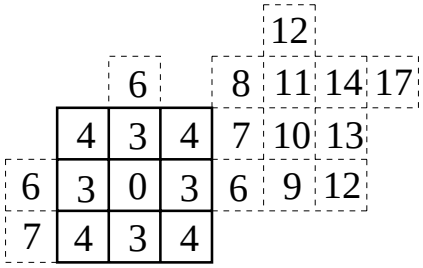


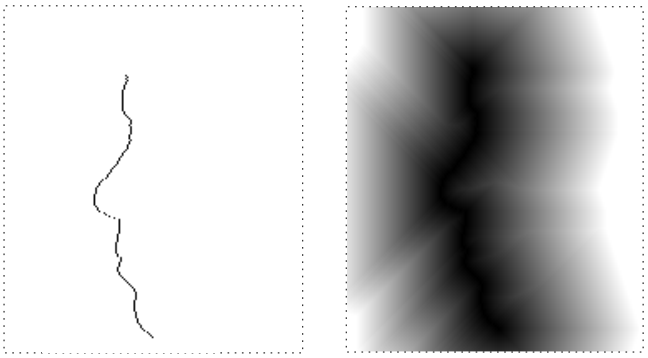
Figure 1.3 - *Calcul d'une distance globale par propagation des distances locales*

L'utilisation d'autres valeurs pour les paramètres $(a, b)/q$ ainsi que de masques de voisinage plus étendus (5×5, 7×7, ...) est discutée dans [12].

1.2.2 Carte de distance

1.2.2.1 Génération

Nous y avons déjà fait allusion brièvement au point précédent, par définition, une *carte de distance* fournit en chacun des points qui la compose, la plus petite distance qui sépare le point considéré d'un point non nul de l'image binaire dont elle est issue. Cette image binaire a pu préalablement être obtenue à partir d'une image numérisée en tenant compte de caractéristiques telles que la localisation de contours, de zones ayant une texture particulière, de points angulaires ou de marqueurs. Dans le cadre de ce travail, nous travaillerons avec le contour du profil tel qu'illustré à la figure 1.4. Cette figure représente une image binaire où seuls les points appartenant au profil ont été noircis. On y a également représenté sa carte de distance, en associant aux différentes distances un niveau de gris particulier (du plus foncé au plus clair, pour des distances croissantes). Un détail de cette carte est repris dans la figure 1.5 qui permet de visualiser le contour du profil en gris, et les distances associées aux éléments d'image voisins (la distance associée aux points appartenant au contour est trivialement nulle).



(a) (b)

Figure 1.4 - *Une image de profil binarisée (a) et la carte de distance qui lui est associée (b)*

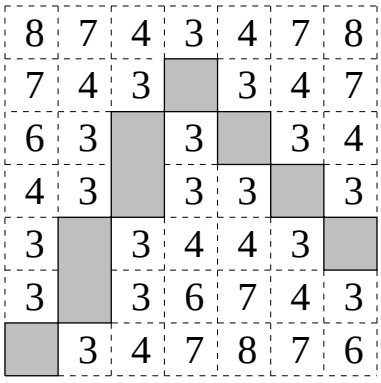


Figure 1.5 - *Carte de distance (détail)*

En adoptant un sens de parcours de gauche à droite et de haut en bas, on remarque que la distance associée à un point situé *en dessous du contour* ne dépend que des points voisins passés et non futurs. Ainsi, si l'on prend l'exemple de la case associée à la distance 6 en avant dernière ligne de la figure 1.5, cette distance est obtenue en prenant le minimum de 3*+4*, 4*+3*, 4*+4* et 3*+3* où les premiers chiffres représentent les distances associées aux cases voisines *passées* et les seconds (en italique) les distances propagées: 3 si on se trouve sur un axe, 4 sur une diagonale. Mathématiquement, cette relation peut s'écrire comme suit:

$$d_{\bigtriangledown}(i,j) = min \left\{ \begin{array}{lll} d(i-1,j-1)+4, & d(i-1,j)+3, & d(i-1,j+1)+4, \\ d(i,j-1)+3 & & \end{array} \right\} \tag{1.2}$$

où $_{\bigtriangledown}$ représente le sens de parcours de gauche à droite et de haut en bas. En inversant celui-ci et en ne considérant que les points situés *au-dessus du contour*, nous pouvons écrire une relation similaire:

$$d^{\bigtriangleup}(i,j) = min \left\{ \begin{array}{lll} & & d(i,j+1)+3, \\ d(i+1,j-1)+4, & d(i+1,j)+3, & d(i+1,j+1)+4 \end{array} \right\} \tag{1.3}$$

où $^{\bigtriangleup}$ représente le sens de parcours inversé. Il en est ainsi du 7 obtenu en première ligne, deuxième case, qui peut être vu comme étant le minimum entre 3*+4*, 4*+3*, 7*+4* et 4*+3*.

On remarquera qu'en appliquant successivement les formules 1.2 et 1.3 sur l'*ensemble* de l'image binaire, et pour autant que celle-ci ait été préalablement initialisée à zéro pour tous les points du contour et l'infini ailleurs, une carte de distance sera automatiquement générée *quelle que soit la forme du contour et sans devoir se soucier de quel côté de la courbe nous nous trouvons.* Tel est l'algorithme du chanfrein, basé sur un double parcours de l'image binaire dont on veut générer la carte de distance. En reprenant les notations utilisées à la section 1.2.1, cet algorithme nécessite $6(kl - N)$ calculs de minima et $8(kl - N)$ sommations. Comme annoncé, la proportionnalité au paramètre N, la longueur du contour, a effectivement bien disparu (complexité d'ordre $O(kl)$).

1.2.2.2 Utilité

Le moyen de générer une carte de distance ayant été précisé, attardons-nous à présent sur son utilité. Une carte de distance offre la possibilité de calculer *rapidement* la distance qui sépare deux contours donnés, par exemple deux profils. Le processus est illustré à la figure 1.6. Cette figure reprend la carte de distance de la figure 1.5 et y superpose un second contour, celui dont on veut connaître la distance par rapport au premier. La distance globale qui sépare les deux courbes (30 dans cet exemple) est donnée par la simple sommation de toutes les distances rencontrées le long du deuxième contour. Une distance moyenne peut être obtenue en normalisant ce résultat par la longueur de ce dernier contour.

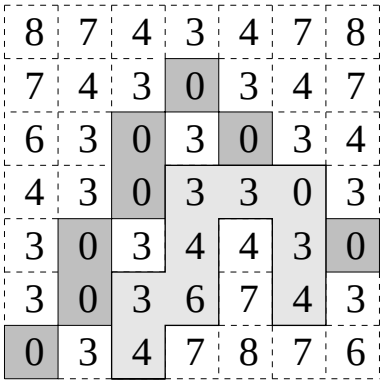


Figure 1.6 - *Distance entre deux contours*

Si jamais la position relative de ces contours se trouvait modifiée, une nouvelle distance globale pourrait être obtenue de la même façon, mais en un temps de calcul négligeable, la carte de distance ayant déjà été générée à l'étape précédente. *Ce gain de temps, lié au calcul préliminaire d'une carte de distance globale, représente l'avantage majeur de la technique proposée ici.*

1.2.3 Mise en correspondance de profils

Intéressons-nous à présent au problème de la mise en correspondance de profils abordé en [50]. On appellera profils de *référence*, les profils repré-

sentatifs de chaque client stockés dans la base de données d'un serveur central, et profil *candidat*, le profil de la personne qui demande l'accès au système. Un tel accès lui sera accordé si la distance entre les profils candidat et de référence est suffisamment faible, c'est-à-dire en dessous d'un seuil d'acceptation fixé. Cette distance est calculée conformément à la technique proposée dans la section précédente.

La première étape consiste donc à générer la carte de distance relative au profil de référence. En superposant le contour candidat sur cette carte, une estimation de la distance qui sépare les deux profils est alors obtenue. Bien sûr, comparer directement ces deux profils n'a que peu de sens si l'on ne tient pas compte des transformations susceptibles d'affecter le profil d'une prise de vue à l'autre. Ces transformations peuvent être tant de nature géométrique, comme celles liées aux variations de la position relative entre le profil et la caméra, que physique, telles les variations engendrées par les parties non rigides du profil. Ces dernières étant difficiles à caractériser, nous nous limiterons à compenser les paramètres liés à la position relative du profil et de la caméra, à savoir:

- une translation (t_x, t_y) dans le plan image (x, y)
- un facteur d'échelle global z
- une rotation θ_{xy} dans le plan image (x, y).

En appliquant ces transformations sur le profil candidat, on obtient un *profil compensé*. Ce profil compensé est à nouveau superposé sur la carte de chanfrein du profil de référence et une nouvelle distance globale est calculée. La meilleure mise en correspondance est obtenue pour le jeu de valeurs de $\{t_x, t_y, z, \theta_{xy}\}$ qui minimise la distance globale entre les deux profils. En d'autres mots, cela revient à minimiser une fonction "distance" qui dépend de quatre paramètres $(t_x, t_y, z, \theta_{xy})$ et dont l'évaluation se fait par la technique du chanfrein. Un tel problème peut être résolu numériquement par n'importe quel algorithme de minimisation de fonctions multidimensionnelles. L'algorithme du *Simplexe de Downhill* fut choisi en raison de sa rapidité d'exécution et sa propriété à ne procéder qu'à des évaluations de fonctions, et non de leurs dérivées [54]. Cet algorithme est détaillé en annexe A. La combinaison des algorithmes du chanfrein et du simplexe fut déjà mise en œuvre précédemment au laboratoire par [61]. Notons que la

normalisation proposée ne permet pas de compenser toutes les transformations tridimensionnelles se rapportant à la position relative du visage et de la caméra.

La figure 1.7 illustre la procédure générale de mise en correspondance du chanfrein dans le cadre de l'authentification de profils. Le profil candidat est placé sur la carte de distance relative au profil de référence, ce qui permet d'obtenir la distance qui sépare ces deux profils bruts. En minimisant cette distance, nous trouvons les paramètres $\{t_x, t_y, z, \theta_{xy}\}$ optimaux qui appliquent au mieux un profil sur l'autre. La distance résiduelle est alors utilisée pour décider si les deux profils appartiennent ou non à la même personne.

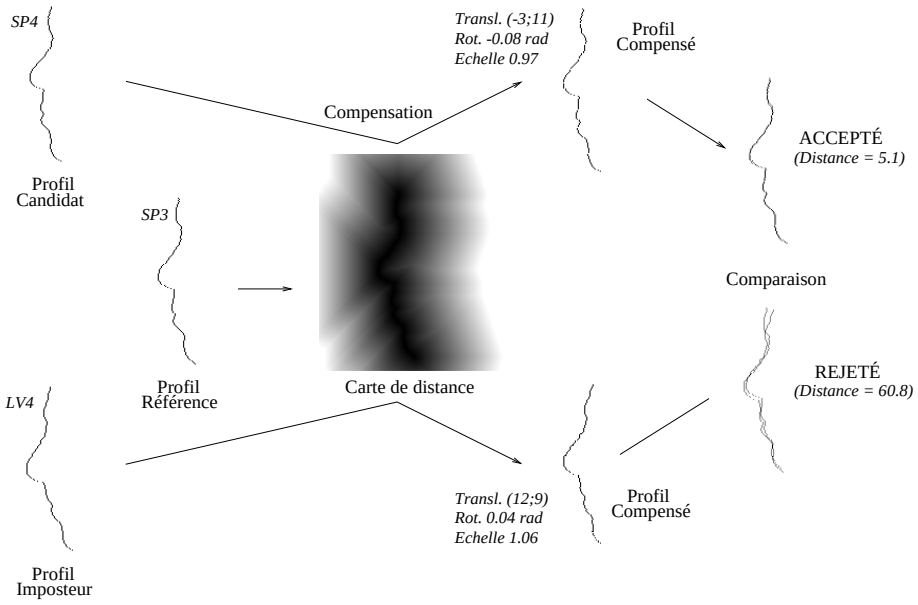


Figure 1.7 - *La procédure de mise en correspondance de profils (exemple réel)*

1.2.4 Optimisations

1.2.4.1 Valeurs initiales

Pour éviter que l'algorithme du simplexe ne se mette à converger vers un minimum local seulement, il faut veiller à ce que les valeurs des paramètres $\{t_x, t_y, z, \theta_{xy}\}$ utilisées pour initialiser l'algorithme soient les plus proches possible de leur valeur optimale. Ainsi, les paramètres $\{t_x, t_y\}$ sont estimés en comparant les positions de l'extrémité du nez entre les profils de référence et candidat. Le paramètre z est approché par le rapport des hauteurs entre profils et θ_{xy} par zéro. Ces valeurs sont utilisées pour initialiser une version basse résolution de l'algorithme du chanfrein/simplexe, où contours de profils et carte de distance ont été sous-échantillonnés par un facteur 4. En sortie, nous obtenons des valeurs raffinées pour $\{t_x, t_y, z\}$ et une bonne estimation de θ_{xy}. Ces valeurs servent enfin à lancer la recherche finale en pleine résolution. Le fractionnement de l'algorithme en différentes résolutions permet non seulement de bénéficier d'un gain de temps appréciable mais aussi de réduire l'influence de minima locaux, moins prononcés dans le domaine sous-échantillonné.

1.2.4.2 Equations de compensation

Une attention particulière doit également être portée aux équations décrivant les transformations géométriques de translation, rotation et facteur d'échelle, appliquées au profil candidat. En particulier, il faut veiller à réduire au maximum l'influence d'un paramètre sur l'autre. Par exemple, lors de la compensation de l'effet de la rotation θ_{xy}, nous aurions pu choisir un centre de rotation[5] éloigné du centre de masse du profil. D'un point de vue théorique, un tel choix n'est pas critique, puisque toute transformation rigide peut être décrite comme résultant d'une rotation et d'une translation, indépendamment du centre de rotation choisi. Néanmoins, en pratique, le choix d'un centre de rotation éloigné du centre de masse de l'objet induit une translation supplémentaire qui doit être compensée ultérieurement. Cette translation, proportionnelle à la distance entre le centre de rotation et l'objet considéré, peut prendre des valeurs considérables. Ainsi, l'algorithme du simplexe éprouvera certaines difficultés lorsque, proche de

5. Souvent associé par défaut aux coordonnées $(0, 0)$

la solution optimale, il lui restera à affiner la valeur de θ_{xy} [6]. En modifiant légèrement θ_{xy}, le profil candidat subira un déplacement additionnel et le vecteur de translation (t_x, t_y) précédemment trouvé ne sera plus correct. Ce problème peut être évité en centrant la rotation et l'homothétie autour du centre de masse du profil (m_x, m_y). Les équations de la transformation appliquée au profil candidat sont alors les suivantes:

$$\begin{aligned} x' &= t_x + m_x + z\{(x - m_x)\cos\theta_{xy} + (y - m_y)\sin\theta_{xy}\} \\ y' &= t_y + m_y + z\{(m_x - x)\sin\theta_{xy} + (y - m_y)\cos\theta_{xy}\} \end{aligned} \tag{1.4}$$

1.2.4.3 Distance quadratique

Afin de pondérer davantage les points du profil candidat qui se trouvent éloignés du profil de référence lors de la minimisation de la distance du chanfrein [7], la distance globale séparant les deux contours est obtenue par sommation quadratique des distances individuelles et non par la simple somme suggérée en fin de section 1.2.2.

1.2.4.4 Double mise en correspondance

Notons enfin que la mise en correspondance proposée n'est pas une opération symétrique et qu'un problème peut se poser lorsque le profil candidat s'étend au-delà du profil de référence [8], comme illustré à la figure 1.8. Dans ce cas, la distance résiduelle obtenue en fin de minimisation ne correspond plus à la meilleure mise en correspondance possible entre les deux profils. En effet, si la mise en correspondance avait été correcte, des points du profil candidat se seraient étendus au-delà du profil de référence et auraient été associés à des distances élevées. *Pour éviter une telle situation, deux distances résiduelles seront calculées: celle du candidat mis en correspondance*

6. Un problème similaire se pose pour le facteur d'échelle z.

7. Les profils candidats et de référence sont sensés appartenir à la même personne. Il nous faut donc pénaliser autant que possible les points du profil candidat n'ayant pu être mis en correspondance avec la référence.

8. Ce cas se présente, par exemple, lorsque l'extraction du profil s'arrête à la hauteur des cheveux et que l'image candidate présente un front plus dégagé que l'image de référence.

avec la référence et celle de la référence mise en correspondance avec le candidat. La distance finale est donnée par la distance résiduelle la plus faible des deux. Cette façon de procéder augmente malheureusement la durée de l'authentification par un facteur deux et empêche le calcul préliminaire des cartes de distance nécessaires à la mise en correspondance des profils. En effet, à l'instar de la carte relative au profil de référence, la carte candidate ne peut être générée avant que le candidat se soit présenté face au système.

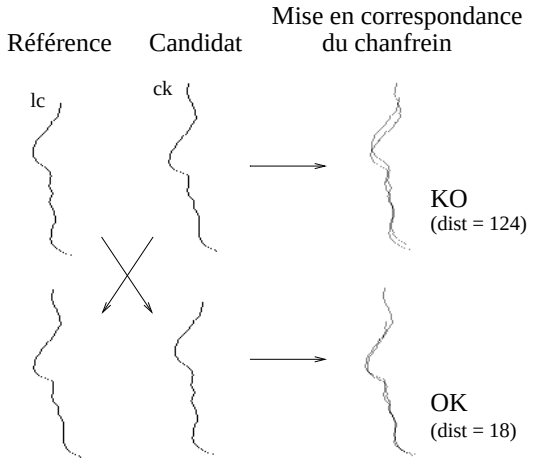


Figure 1.8 - *Asymétrie de la mise en correspondance du chanfrein*

1.3 Autres pistes étudiées

Parallèlement à la modalité traitant du relief du profil (section 1.2), d'autres pistes de recherche liées elles aussi au profil ont été suivies. Toutes sont présentées dans cette section, mais offrent en général des performances médiocres. Seule une modalité traitant les niveaux de gris le long du profil a pu être validée. Celle-ci sera combinée par la suite à la modalité "relief du profil" pour donner naissance à *l'expert profil*, introduit au chapitre 4.

1.3.1 Extrapolation du profil à partir de la vue de face (échec)

L'analyse de la vue de profil n'est pas aussi simple à mettre en œuvre que celle de la vue de face, lors de l'implémentation pratique d'un système d'authentification. Ainsi, la vue de profil nécessite le recours à une caméra perpendiculaire à la direction du regard de la personne qui désire s'authentifier, personne qui pourra de ce fait éprouver certaines difficultés à se positionner correctement.

Afin de palier cet inconvénient, nous nous sommes intéressés à la synthèse d'un profil à partir d'une image frontale. Par analogie à la méthode proposée en [32], le *pseudo-profil*[9] est obtenu en sommant les valeurs de luminance le long des lignes qui composent l'image du visage de face. Un visage et son pseudo-profil sont illustrés à la figure 1.9.

Figure 1.9 - *Un visage et son pseudo-profil*

Alors que dans [32], le relief du pseudo-profil est traité dans sa globalité, il s'est avéré impossible d'en faire de même, ce relief s'étant révélé trop sensible aux conditions d'éclairage et à la présence d'ombres. Quelques points caractéristiques seulement ont pu être localisés de façon automatique sur les visages utilisés lors des tests. Il s'agit de l'ordonnée des sourcils, des pupilles, des narines et de la commissure des lèvres (figure 1.9). Pour réduire

9. On utilisera ce qualificatif pour distinguer ce profil du profil vrai traité dans le cadre du chapitre 1.2.

l'effet du bruit sur la localisation de ces différents points, la recherche est effectuée sur une image préalablement filtrée passe-bas.

Le tableau 1.1 donne les distances sourcils-pupilles et yeux-narines normalisées par rapport à la distance yeux/bouche pour 4 visages extraits de la base de données M2VTS[10] et 4 prises de vues différentes. Une astérisque indique une caractéristique n'ayant pu être localisée, deux astérisques, une localisation erronée.

	SP	PF	PP	VW
Prise de vue 1	0.18-0.63	0.16-0.62	0.22-(*)	0.16-0.64
Prise de vue 2	0.13-0.51	0.17-0.53	0.22-(*)	(**)
Prise de vue 3	0.16-0.51	0.18-0.51	0.24-0.68	0.20-0.61
Prise de vue 4	0.18-0.60	0.20-0.57	0.17-(*)	0.19-0.61

Tableau 1.1 - *Distances sourcils-pupilles et yeux-narines normalisées par rapport à la distance yeux-bouche, pour 4 personnes de la base de données M2VTS. (*) caractéristique n'ayant pu être localisée (**) localisation erronée.*

Ce tableau nous montre que la variation des distances entre les différentes prises de vues d'un même visage est approximativement égale à la variation observée entre différents visages. Ce résultat a pu être confirmé pour l'ensemble des visages appartenant à la base de données de test. Nous pouvons ainsi conclure que même les caractéristiques les plus faciles à localiser sur un pseudo-profil, ne peuvent être extraites de façon fiable. Comme mentionné précédemment, la cause de ce mauvais résultat est principalement due à la grande sensibilité du pseudo-profil aux conditions d'éclairage et particulièrement à l'orientation du visage par rapport aux sources lumineuses. Dès lors, les caractéristiques issues d'un pseudo-profil à partir de la vue de face, n'ont pu être utilisées par la suite.

1.3.2 Profil intermédiaire (échec)

L'utilisation de contours du visage issus de vues différentes de celle du profil à 90^o permet de compléter l'information relative au profil utilisée jusqu'à présent. Ces vues intermédiaires doivent être sélectionnées en tenant compte

10. Voir chapitre 3.

(a) (b)

Figure 1.10 - *Vue de profil à* 90^o *(a) et profil intermédiaire (b)*

des remarques suivantes:

- pour éviter l'influence des cheveux, masquant le contour du visage que l'on désire extraire, la vue considérée doit être suffisamment éloignée de la vue de face pour laquelle la présence de cheveux empêche toute extraction du contour extérieur du visage;

- la vue considérée doit être suffisamment éloignée de la vue du profil afin d'accéder à des caractéristiques autres que celles déjà présentes dans le profil à 90^o;

- un même angle de vue doit pouvoir être sélectionné quelque soit la personne considérée.

Selon ces contraintes, une seule vue intermédiaire a pu être retenue. Elle correspond à l'image sur laquelle le *nez apparaît tangent au visage*, comme illustré à la figure 1.10. Nous appellerons *profil intermédiaire*, le relief du contour du visage relatif à cette vue.

Malheureusement, les résultats d'une authentification basée sur le relief du profil intermédiaire seul, se sont avérés mauvais et cette méthode fut à son

tour rejetée[11]. Les raisons en sont les suivantes:

- le profil extrait de la vue intermédiaire ne présente aucune caractéristique marquante et est de ce fait fort semblable d'une personne à l'autre[12];

- il est difficile de se départir complètement de l'influence des cheveux. Même lorsque le profil intermédiaire est limité en deçà des sourcils, des cheveux longs sont susceptibles de se détacher de l'arrière-plan et poser problème lors de l'extraction de contour (voir le cas ID4 à la figure 1.11). La vue de profil échappe à ce problème, les cheveux gênants étant généralement masqués par le visage.

1.3.3 Profils combinés (échec)

Comme le profil intermédiaire ne peut à lui seul apporter de nouvelles caractéristiques pertinentes, nous avons essayé de combiner les reliefs du profil à 90^o et intermédiaire au sein d'une même image binaire, le profil intermédiaire étant restreint à la zone située entre la bouche et les sourcils comme illustré à la figure 1.11. Ces images sont alors traitées par l'algorithme de mise en correspondance du chanfrein afin de trouver la meilleure mise en correspondance possible entre les profils combinés candidat et référence.

Dans l'image combinée, les deux profils sont placés à l'endroit même où ceux-ci sont localisés dans leurs vues respectives. Nous accédons ainsi à une nouvelle caractéristique biométrique liée à la façon dont la tête est capable de pivoter autour de son axe. Malheureusement, la disposition relative du profil vrai et du profil intermédiaire est trop variable d'une prise de vue à l'autre pour que la méthode proposée soit capable d'offrir de bons résultats[13]. Ces piètres résultats peuvent aussi s'expliquer par la mauvaise qualité de la segmentation du profil intermédiaire, principalement influencée par la présence de cheveux longs comme dans le cas ID4 de la figure 1.11. L'idée des profils combinés fut donc rejetée à son tour.

11. Le profil intermédiaire offre, en terme de fausse acceptation et faux rejet (voir section 4.3), des performances de près de 3 fois moins bonnes que le profil vrai.

12. Ce qui se traduit par un taux de fausse acceptation élevé.

13. A nouveau, une dégradation des performances du système par un facteur 3 est à noter, par rapport au cas du profil à 90^o seul.

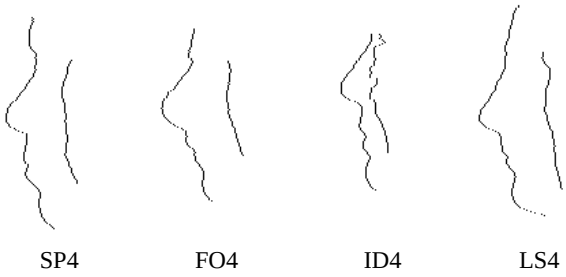


Figure 1.11 - *Exemples de profils combinés*

1.3.4 Corrélation des niveaux de gris

Une dernière piste étudiée en vue d'améliorer les performances de la mise en correspondance du chanfrein, consiste à comparer les niveaux de gris (ou luminance) de l'image candidate avec ceux de l'image de référence. Après avoir trouvé les meilleurs paramètres de compensation $\{t_x, t_y, z, \theta_{xy}\}$ lors de la mise en correspondance du chanfrein, ces mêmes paramètres sont appliqués sur l'image de profil du candidat pour l'amener en correspondance avec l'image de référence. Il devient alors possible de comparer chaque point de l'image candidate avec son homologue de l'image de référence. Néanmoins, nous veillerons à normaliser les niveaux de gris des deux images pour se départir partiellement des effets liés à un éventuel changement d'éclairage global entre les deux prises de vues. Deux types de normalisation de la luminance ont été étudiés:

- une *normalisation de contraste (essai infructueux)* qui étend la distribution des niveaux de gris de chaque image jusqu'à l'obtention de la dynamique maximale[14], soit de 0 à 255 pour une représentation sur 8 bits. Malheureusement, cette manière de procéder pose problème lorsque des zones très claires sont présentes dans une image mais pas dans l'autre[15]. Malgré des conditions d'éclairage identiques pour les deux images, l'une d'elle sera proche de la dynamique maximum, tandis que l'autre sera soumise à l'opération de normalisation.

14. Cette normalisation produit un effet similaire à ce que nous aurions pu obtenir en choisissant de normaliser la variance de la luminance.

15. Par exemple, lorsque la bouche est ouverte et que la dentition est visible.

Il en résultera une modification globale de la luminance de l'une des deux images seulement, ce qui causera l'échec de la comparaison des niveaux de gris.

– une *normalisation de la luminance moyenne (essai retenu)* qui fixe la valeur moyenne de la luminance des deux images au milieu de la dynamique maximale (soit 127), en décalant la luminance de chaque pixel. Cette méthode fut adoptée pour la suite de ce travail.

Des méthodes plus complexes permettent de se départir de l'effet d'un éclairage de biais et des ombres projetées sur une moitié du visage[16]. Il s'agit principalement de méthodes d'égalisation de luminance basées sur le calcul d'histogrammes partiels, comme détaillé dans [47].

La mesure de correspondance entre niveaux de gris est obtenue par un critère d'*Erreur Quadratique Moyenne* (EQM), calculé sur un ensemble de pixels bordant le profil, comme illustré à la figure 1.12. Afin de se départir au mieux de l'influence néfaste des cheveux, seule une bande de 25 pixels localisée le long du contour du profil a été considérée. Au-delà de 25 pixels, les performances d'authentification se détériorent rapidement. Aussi, la bande sur laquelle l'EQM est calculée a préalablement été filtrée passe-bas afin de lisser les zones du visage sujettes à des variations assimilables à du bruit haute fréquence, comme la zone des sourcils par exemple, et qui détériorent sensiblement la valeur de l'EQM.

Cette technique offre des performances similaires à celles offertes par la modalité liée au contour du profil (voir sections 4.4 et 4.5). Dans le chapitre 4, ces deux modalités seront combinées pour former un module spécialisé dans l'authentification de la vue de profil. Cet *expert profil* bénéficiera de performances accrues par rapport à chacune des modalités profils prises séparément.

16. De telles images ne sont pas présentes dans la base de données utilisée.

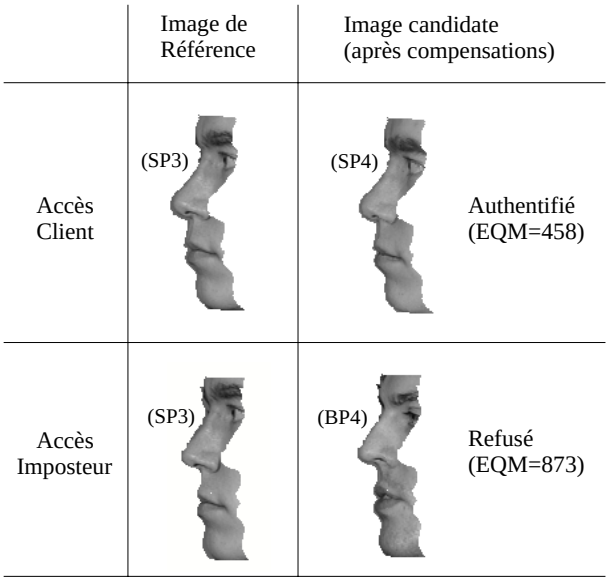


Figure 1.12 - *Authentification par corrélation de niveaux de gris le long du profil*

1.4 Segmentation et extraction du profil

Jusqu'à présent, nous supposions que le contour du profil avait pu être extrait de l'image relative à la vue de profil. Dans cette section, nous verrons comment y parvenir.

La technique mise en œuvre se base sur l'hypothèse d'un *fond uniforme dont la teinte diffère de celle du visage*. Cette teinte peut néanmoins varier selon les conditions lumineuses. En outre, le fond doit présenter *une surface visible plus étendue que n'importe quelle autre surface de couleur uniforme présente dans l'image* (ce qui s'avère être souvent le cas en pratique). Dans le cas d'un fond non uniforme, mais supposé néanmoins fixe, une technique basée sur la soustraction de l'image à traiter avec une image où seul le fond est présent peut être mise en œuvre [63].

La segmentation du profil est divisée en deux étapes: la première dissocie le visage de profil du fond tandis que la seconde traite de l'extraction et de la normalisation du profil proprement dit.

1.4.1 Segmentation du visage par technique de Regroupements de Couleurs[17]

Si le contraste entre le fond homogène et le visage est suffisant, diverses techniques simples d'extraction de contours peuvent être mises en œuvre pour segmenter le profil à partir d'une image telle que celle représentée à la figure 1.10 (a). Ces techniques se basent généralement sur la différence de luminosité entre pixels voisins, ce qui revient encore à évaluer une dérivée spatiale de la luminance sous la forme d'un gradient, laplacien, ou tout autre opérateur qui réalise l'implémentation particulière d'un filtre passe-haut [40, 57]. Ces techniques seront généralement utilisées dans le cas d'images noir et blanc.

Par ailleurs, une image couleur peut très bien contenir des transitions (contours) caractérisées par une différence de *teinte* uniquement. Ainsi, les images utilisées dans le cadre de ce travail (voir chapitre 3) n'offrent qu'un contraste très faible entre la luminance du fond et celle de la peau. C'est donc sur l'information de couleur que nous devrons travailler. La figure 1.13 nous montre une telle image et le détail d'une zone particulière du profil. L'agrandissement est tel que l'on distingue clairement les éléments de l'image (*pixels*) et combien il est difficile de pouvoir dissocier le visage du fond gris. Le contraste est à ce point faible que le contour recherché est masqué par le bruit engendré par la caméra, visualisé sous la forme de pixels dont la luminance diffère fort de celle des voisins.

Remarquons néanmoins que si le problème se pose au niveau local (l'agrandissement de la figure 1.13), il n'en va pas de même au niveau global (l'image entière) où le contour du profil se distingue relativement bien du fond. Cette constatation plaide en faveur de l'utilisation d'une méthode de segmentation traitant l'image dans sa totalité.

Ces deux remarques conduisent à la technique de segmentation par regroupements de couleurs, proposée par [43] et adoptée dans le cadre de ce travail.

Cette segmentation peut être résumée comme suit. Premièrement, l'image de départ est filtrée passe-bas afin de lisser les différentes composantes de couleur et de réduire l'effet du bruit de la caméra au sein de chacune des composantes. Ensuite, un histogramme[18] bidimensionnel est calculé sur

17. *Color clustering*, en anglais.

18. Un histogramme associe à chaque valeur que peut prendre une variable donnée (ici,

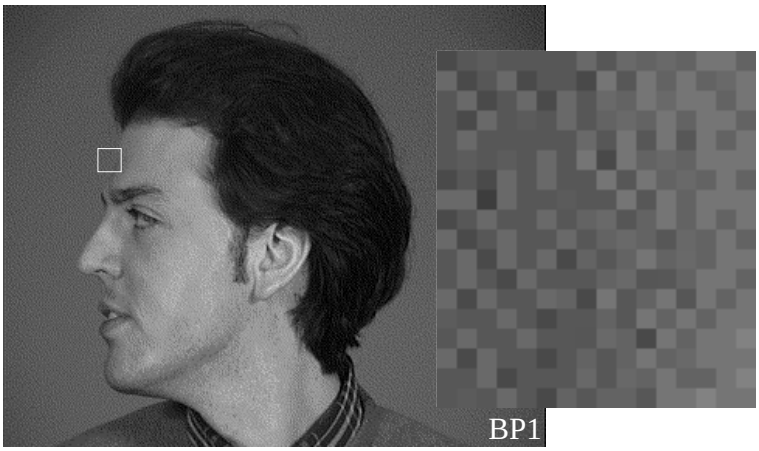


Figure 1.13 - *Détail d'une image de profil*

base des deux composantes de couleur qui distinguent au mieux le visage et le fond gris, soit le rouge (R) et le vert (V) pour une décomposition RVB[19]. Ces composantes sont préalablement normalisées pour se départir de l'influence de la luminosité:

$$\begin{aligned} r &= R/(R+V+B) \\ v &= V/(R+V+B) \end{aligned}$$

où r et v représentent les composantes rouge et verte normalisées et R, V et B, les trois couleurs primaires.

Un tel histogramme est illustré à la figure 1.14. Chaque cellule $c_{(x,y)}$ du plan (r, v) fournit le nombre de pixels ayant comme composantes de couleur (après quantification), les composantes (x, y) qui caractérisent la cellule envisagée.

Un lien entre chaque cellule de l'histogramme est établi en faisant pointer chacune d'elles vers la plus grande de ses voisines (la taille des cellules et

les pixels d'une image) le nombre de ses occurrences.

19. Une décomposition en composantes principales aurait permis de trouver les couleurs primaires qui caractérisent au mieux la teinte de la peau. Nous nous sommes néanmoins limités à la décomposition classique en composantes rouge-vert-bleu.

du voisinage sont des paramètres de l'algorithme). La figure 1.14 illustre le cas où seuls les 4 voisins axiaux sont pris en considération. Comme le fond de l'image représente la plus grande surface uniforme dans la vue de profil, sa localisation principale au sein de l'histogramme est donnée par la cellule totalisant le score le plus élevé. En regroupant toutes les cellules qui pointent directement ou indirectement vers celle-ci, et en les projetant à nouveau dans le domaine image, le fond est isolé du visage. Un post-traitement est enfin nécessaire pour éviter que des pixels de l'image ne se retrouvent isolés au milieu d'une zone déterminée. Ainsi, un pixel n'ayant pu être classé comme faisant partie du fond, mais dont les voisins le sont, s'y trouvera inclus pour autant que la cellule de l'histogramme dont il est issu soit adjacente à l'une des cellules qui caractérisent le fond.

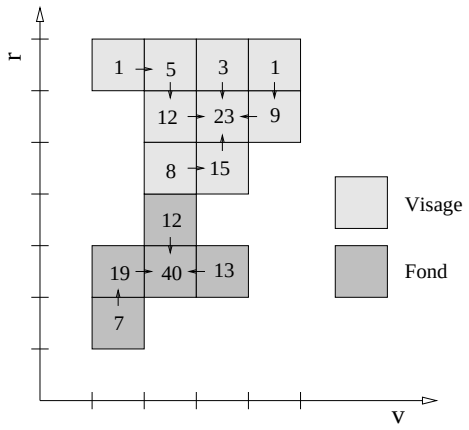


Figure 1.14 - *Segmentation par regroupements de couleurs*

En discrétisant les composantes r et v en 100 pas (histogramme à 10,000 cellules) et en restreignant le critère de voisinage aux cellules adjacentes (voisins immédiats), cette technique débouche sur les excellents résultats illustrés à la figure 1.15.

1.4.2 Extraction de profils normalisés

Une fois le visage segmenté du fond de l'image, il reste à en extraire le contour profil. Afin que la technique de mise en correspondance du chanfrein puisse donner de bons résultats, cette extraction doit se limiter aux parties

(a)

(b)

Figure 1.15 - *Deux exemples de segmentation du visage par technique de regroupements de couleurs (fond uniforme)*

invariantes du profil. Il faut donc:

- exclure le front lorsque celui-ci est susceptible d'être recouvert par des cheveux tombants;
- exclure la partie inférieure du profil constituée du dessous du menton et du cou, le relief de cette zone étant particulièrement sensible à l'inclinaison de la tête;
- exclure la partie inférieure du profil pour des personnes portant la barbe ou la moustache.

Pour pouvoir tenir compte de la topologie du visage (coupe de cheveux, présence d'une moustache, barbe, etc.), on donne au gestionnaire de la base de données client le choix de différents modes d'extraction du profil lors de la définition d'un nouvel utilisateur[20]. Une fois le mode choisi, il sera utilisé *à chaque fois* que l'utilisateur (ou un imposteur usurpant son identité) se présentera devant le système d'authentification. Un premier mode segmente le profil entier et suppose la présence de cheveux courts et l'absence de barbe. Un deuxième mode ne sélectionne que la partie inférieure du profil et doit être utilisé lorsque des cheveux longs sont susceptibles de recouvrir le front. Un troisième mode segmente la partie supérieure du profil uniquement et convient pour les utilisateurs portant la barbe ou la moustache. Enfin, le quatrième mode résulte de la combinaison des deux modes précédents et ne sélectionne que la partie centrale du profil.

Il est également essentiel qu'au sein de chaque mode, la même partie du profil soit extraite indépendamment de la prise de vue, pour un utilisateur donné. Si les profils mis en correspondance ne sont pas proprement calibrés, un biais affectera la distance du chanfrein résiduelle de façon similaire au problème décrit en fin de section 1.2.4 (figure 1.8).

Tenant compte de ces remarques, l'extraction de profil a été décomposée en différentes étapes (voir figure 1.16):

- *Localisation de l'extrémité du nez.* Les coordonnées correspondantes à l'extrémité du nez $(n_x^b; n_y^b)$ sont obtenues en recherchant le point du profil le plus proche du bord de l'image. Ce point peut néanmoins

20. Le mode est donc choisi en fonction de l'apparence de la personne au moment de l'apprentissage.

se retrouver dans les cheveux lorsque le visage est incliné vers le bas. Dès lors, la recherche de celui-ci est limitée au sein d'une fenêtre horizontale couvrant la zone centrale du visage uniquement.

- *Localisation du dessus du nez.* Les coordonnées correspondantes au dessus du nez $(n_x^h; n_y^h)$ sont données par le premier extremum local rencontré lorsqu'on parcourt le profil depuis l'extrémité du nez vers le front [21].

- *Segmentation brute du profil.* Les deux points précédemment trouvés nous permettent de calculer la hauteur du nez, $h = n_y^h - n_y^b$. Cette hauteur est utilisée comme distance de référence lors de l'extraction des profils normalisés correspondants aux différents modes décrits plus haut. Ces profils s'étendent sur les intervalles:

  - $[n_y^b - 2h, n_y^h + 0.7h]$ pour le premier mode (profil entier)
  - $[n_y^b - 2h, n_y^h]$ pour le second mode (profil inférieur)
  - $[n_y^b - 0.3h, n_y^h + 0.7h]$ pour le troisième mode (profil supérieur)
  - $[n_y^b - 0.3h, n_y^h]$ pour le quatrième mode (profil central)

- *Corrections apportées à la segmentation brute.* Pour certains visages, le menton est situé au-dessus de la coordonnée $n_y^b - 2h$. Dans ce cas, le dessous du menton et parfois une partie de l'épaule, sont inclus dans le profil segmenté. La partie inférieure du menton peut être aisément localisée grâce à son orientation quasiment horizontale et son gradient vertical élevé. L'épaule, dont la projection recouvre parfois le contour du menton, est détectée par son orientation particulière, une diagonale montante pour un profil dirigé vers la gauche. Comme l'orientation relative du menton et de l'épaule peut fortement varier d'une prise de vue à l'autre, tous les points du contour appartenant à ces deux régions sont alors supprimés du profil normalisé.

Le figure 1.16 illustre ces différentes étapes pour les deux profils de la figure 1.15.

Enfin, pour ne pas entraver l'extraction du profil, il est demandé à l'utilisateur se présentant devant le système, de retirer ses lunettes le temps de l'authentification.

21. Cet extrémum est déterminé par un changement de signe de la dérivée première du relief du profil. Cette dérivée doit être calculée sur une fenêtre suffisamment large pour être insensible au bruit affectant le contour du profil.

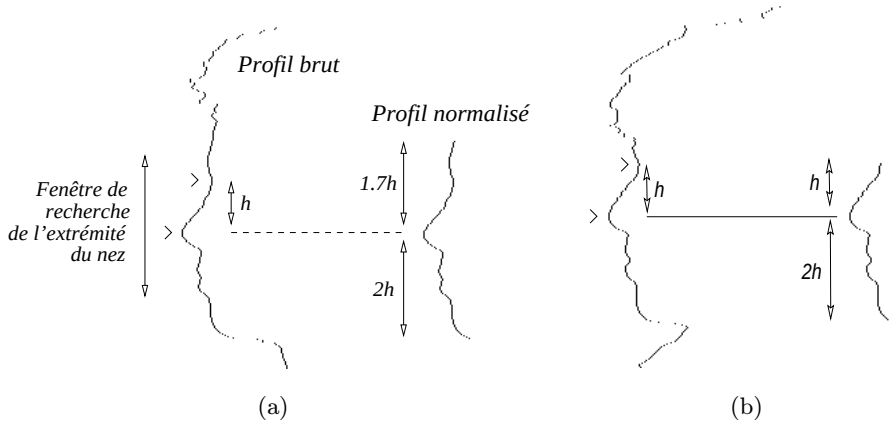


Figure 1.16 - *Extraction de profils normalisés : (a) Premier mode (b) Deuxième mode.*

Chapitre 2

La modalité frontale

2.1 Contexte

Tout comme dans le cas de la vue de profil, les techniques utilisées pour reconnaître ou authentifier un visage vu de face peuvent être classées en deux catégories principales, la première étant basée sur la localisation de caractéristiques biométriques du visage (position des yeux, du nez, de la bouche, hauteur et largeur du visage, etc.), la seconde faisant directement usage des valeurs des niveaux de luminance de l'image correspondante. Une étude comparant ces deux méthodes a été réalisée dans [11].

Au sein de la première catégorie, nous pouvons citer les travaux de Bledsoe [5] qui fut le premier à traiter les caractéristiques géométriques du visage de façon semi-automatisée; l'ordinateur classait des visages sur base de points caractéristiques extraits à la main sur des photographies de visage. Sakai et al [58] ont alors rapidement proposé une méthode automatique pour localiser les diverses caractéristiques utiles du visage. Leur technique se base sur la mise en correspondance d'une image moyenne de la caractéristique à localiser (modèle) avec l'image d'entrée préalablement filtrée. Signalons aussi les travaux de Kanade [34], considéré comme l'un des pionniers dans le domaine de la reconnaissance automatique de personnes. Il définit en particulier 16 rapports de longueurs ou angles particuliers qui discriminent au mieux des visages différents. Plus récemment, le problème de la localisation automatique de caractéristiques liées au visage fut investigué par [16] et [10].

En ce qui concerne la deuxième catégorie, une variété bien plus grande d'algorithmes est reprise dans la littérature. Les pistes de recherche principales sont mentionnés ci-dessous:

- *Reconnaissance/authentification par mesure de corrélation de luminance.* Dans sa version la plus simple, les niveaux de gris de l'image du visage candidat sont comparés avec ceux de l'image de référence. De nombreuses améliorations sont possibles: pré-traitement de l'image d'entrée, fractionnement de l'image modèle en un ensemble d'images de taille réduite, compensation du mouvement rigide et éventuellement des mouvements élastiques du visage, etc. [69, 11, 39, 53]

- *Utilisation de visages propres.* Par analogie aux vecteurs propres issus d'une décomposition de Karhunen-Loève, un visage est décomposé en une somme pondérée de visages propres. L'authentification ou la classification d'un visage s'effectue alors par l'intermédiaire d'un nombre limité de ces coefficients, résultant en un temps d'authentification ou de recherche réduit. [35, 45, 44]

- *Corrélation de grilles.* Une grille rectangulaire est appliquée sur l'image du visage de référence. En chacun de ses nœuds, un jeu de coefficients représentant le contenu fréquentiel local de l'image est calculé. Une grille identique est alors superposée sur l'image candidate et déformée de façon globale, puis locale afin d'obtenir un jeu de coefficients candidats aussi proche que possible du jeu de référence. Si ces jeux sont suffisamment proches l'un de l'autre, le candidat est authentifié. [37, 67, 38, 20]

2.2 Méthode choisie: corrélation des niveaux de gris

Pour des raisons pratiques de programmation, la méthode choisie est semblable à celle mise en oeuvre à la section 1.3.4 dans le cadre de l'analyse du profil, à savoir basée sur le calcul de la corrélation des niveaux de gris entre une image candidate et une image de référence. Cette technique s'apparente donc aux techniques de corrélation de luminance brièvement décrites ci-dessus. Elle fut initialement étudiée à l'UCL par [6] puis optimisée par [53].

La même normalisation de luminance que celle utilisée dans la section 1.3.4 est appliquée sur les images référence et candidate pour compenser les variations d'éclairage entre prises de vues (normalisation de la moyenne). A nouveau, les images sont filtrées passe-bas afin d'améliorer la mise en correspondance des zones sujettes à des variations assimilables à du bruit haute fréquence (principalement les sourcils) et le même critère EQM est utilisé pour calculer la distance entre les deux images mises en correspondance.

Pour se départir au maximum des variations intrinsèques du visage, seule une fenêtre rectangulaire centrée autour des caractéristiques les plus stables a été utilisée. Cette fenêtre, appelée *fenêtre active*, regroupe les caractéristiques biométriques liées aux yeux, aux sourcils et au nez. Une étude réalisée par [11] confirme ce choix particulier et désigne les yeux et le nez comme étant les caractéristiques les plus discriminantes du visage. Contrairement à [11] qui fait usage de deux fenêtres distinctes pour les yeux et le nez, la méthode proposée ici regroupe ces caractéristiques au sein d'une même fenêtre active, cette combinaison s'étant avérée plus discriminante encore. De telles fenêtres sont illustrés dans la figure 2.1 et sont extraites automatiquement des images d'entrée par l'algorithme décrit dans la section 2.3.

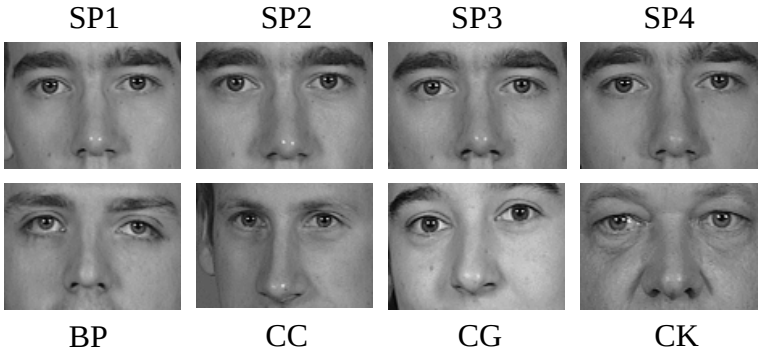


Figure 2.1 - *Exemples de fenêtres utilisées lors de la mise en correspondance de visages vus de face*

Contrairement à la modalité équivalente de la vue de profil qui utilisait les paramètres de compensation issus de la mise en correspondance du chanfrein, nous ne connaissons a priori la transformation géométrique à appliquer pour amener l'image candidate en concordance avec l'image de référence. Aussi, faisons-nous à nouveau usage de l'algorithme du simplexe

(voir annexe A) pour trouver les paramètres t_x, t_y, z et θ_{xy} qui minimisent l'EQM entre les niveaux de gris des deux images. Afin d'accélérer la convergence de l'algorithme, mais aussi pour éviter de converger vers un minimum local, des valeurs approximatives pour $\{t_x, t_y, z, \theta_{xy}\}$ sont utilisées lors de l'initialisation du simplexe. Les valeurs initiales de (t_x, t_y) sont obtenues en comparant la position des yeux dans l'image candidate et l'image de référence tandis que les paramètres z et θ_{xy} sont initialisés à 1 et 0 respectivement.

Jusqu'à présent, aucun des paramètres précités ne permet de corriger une rotation de la tête dans un plan différent de celui du plan image π_{xy}. Pour pouvoir corriger un tel mouvement, il faudrait théoriquement disposer d'un modèle tridimensionnel du visage. Un tel modèle peut être obtenu à partir d'une séquence d'images dans laquelle apparaît le visage en mouvement comme étudié en [55]. Malheureusement, cette opération est coûteuse en temps de calcul et la précision offerte n'est pas suffisante pour pouvoir l'appliquer à l'authentification de visages. Dans le cas qui nous préoccupe, les rotations parasites restent cependant faibles puisqu'il est demandé à la personne de regarder dans la direction de la caméra lors du processus d'authentification. Si le visage avait été une surface plane, ces altérations auraient pu être simplement corrigées par un facteur d'échelle différent dans chaque direction x et y. En supposant la caméra suffisamment éloignée du visage et les rotations du visage suffisamment faibles, l'approximation d'un visage plan donne des résultats acceptables et *nous travaillerons dorénavant avec deux facteurs d'échelle distincts z_x et z_y.*

Cette méthode du double facteur d'échelle permet de réduire sensiblement l'EQM entre deux images de la même personne mais dont l'une est légèrement de biais. Cependant, l'emploi de ces deux facteurs permet aussi de déformer l'image d'un imposteur pour qu'elle puisse mieux ressembler à l'utilisateur qu'il prétend être. Dès lors, il faut veiller à introduire dans la fonction de coût minimisée par l'algorithme du simplexe, un facteur supplémentaire proportionnel à la différence $|z_x - z_y|$. Ainsi, l'algorithme n'utilisera deux facteurs d'échelle différents que si la distance entre les images candidate et de référence est sensiblement améliorée, ce qui sera généralement le cas lorsqu'il s'agit de deux images de la même personne.

Les nouvelles équations de compensation appliquées à la fenêtre candidate sont données par:

$$\begin{aligned} x' &= t_x + m_x + z_x\{(x - m_x)\cos\theta_{xy} + (y - m_y)\sin\theta_{xy}\} \\ y' &= t_y + m_y + z_y\{(m_x - x)\sin\theta_{xy} + (y - m_y)\cos\theta_{xy}\} \end{aligned} \quad (2.1)$$

où (m_x, m_y) désigne les coordonnées du centre de la fenêtre de travail. La fonction de coût peut, quant à elle, être notée sous la forme:

$$EQM\{reference, candidat(t_x, t_y, \theta, z_x, z_y)\} + \alpha|z_x - z_y| \quad (2.2)$$

où le facteur α est fixé manuellement pour n'autoriser qu'une différence maximale de 10% environ entre les deux facteurs d'échelle.

Les figures 2.2 et 2.3 illustrent la technique proposée (cas d'un accès client et d'un accès imposteur respectivement). On y montre l'extraction de la fenêtre active, la recherche de la meilleure mise en correspondance possible entre la fenêtre active et l'image de référence, la superposition de la fenêtre active sur l'image de référence ainsi que le calcul de l'erreur résiduelle.

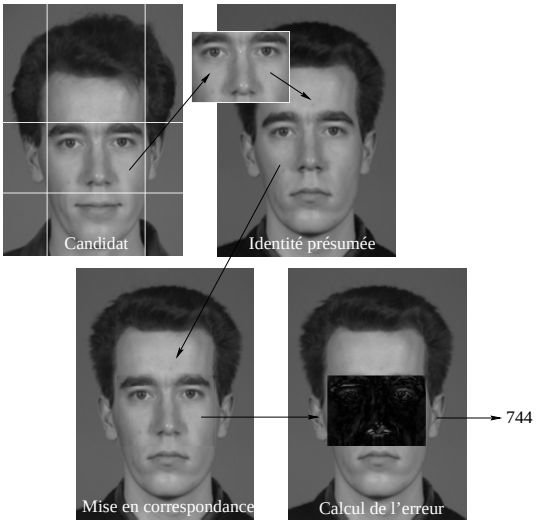


Figure 2.2 - *Authentification frontale: exemple d'accès client*

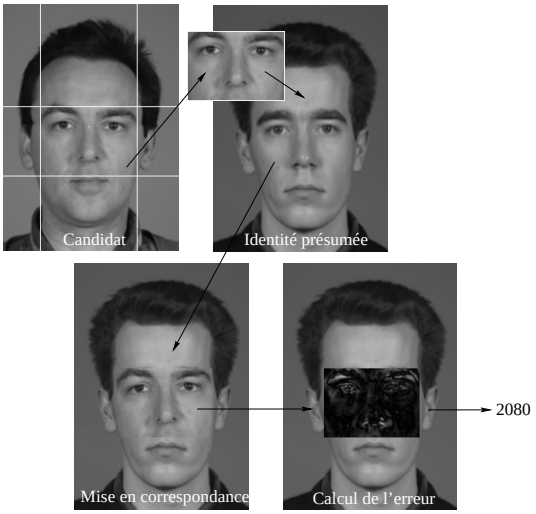


Figure 2.3 - *Authentification frontale: exemple d'accès imposteur*

2.3 Localisation automatique de la fenêtre active

Les fenêtres actives utilisées lors de l'authentification du visage de face sont automatiquement extraites des images frontales[1] par une technique basée sur des projections[2] de gradients, similaire à celle proposée en [11] (voir figure 2.4):

- la projection horizontale du gradient horizontal (PHGH)[3] offre un pic maximal dans l'alignement des yeux, et permet donc de localiser ceux-ci verticalement.
- une bande d'une hauteur comparable à celle des yeux et qui couvre la largeur de l'image est alors centrée autour de l'ordonnée trouvée au

1. Cette opération est réalisée en amont du filtrage passe-bas.

2. Nous appellerons "projection" l'opération qui consite à sommer les valeurs de luminance présentes dans une image, soit le long des lignes qui la composent (projection horizontale), soit le long de ses colonnes (projection verticale).

3. Le gradient horizontal met en évidence les caractéristiques du visage ayant principalement une orientation verticale. Ainsi disparaissent tous les traits horizontaux, comme la bouche et les sourcils de même que les autres traits susceptibles d'engendrer des pics indésirés dans la projection horizontale du gradient.

point précédent. On y calcule la projection verticale du gradient vertical (PVGV)[4] qui offre deux maxima distincts aux abscisses relatives aux yeux et permet de localiser ceux-ci horizontalement.

– Ces trois mesures nous permettent de calculer les coordonnées du point situé au milieu des deux yeux et qui sert à positionner la fenêtre active (une fenêtre dont la taille est fixée à 110×80 pour un format d'image 350×286, voir chapitre 3).

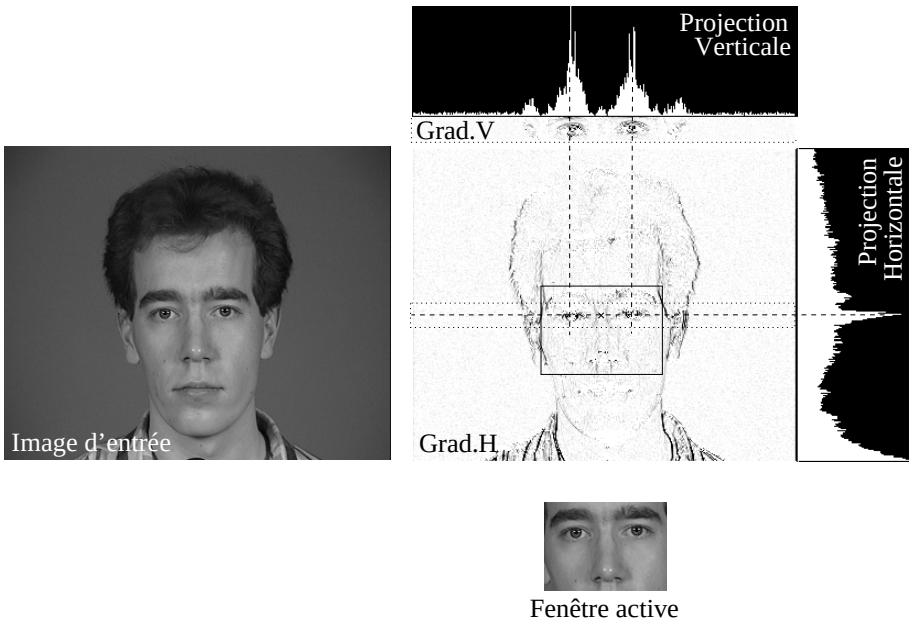


Figure 2.4 - *Extraction automatique de la fenêtre active*

Cette procédure, illustrée à la figure 2.4, donne d'excellents résultats sur l'ensemble de la base de données M2VTS utilisée dans ce travail. Elle est néanmoins mise à défaut lorsque l'utilisateur oublie de retirer ses lunettes, ou si les images sur lesquelles on travaille sont de moins bonne qualité que celles utilisées jusqu'à présent[5]. Pour rendre la méthode plus robuste aux

4. Le gradient vertical mettant en évidence les structures horizontales, on se départit ainsi de l'influence du nez (principalement vertical) dans la projection verticale du gradient.

5. Remarquons la qualité avec laquelle les yeux se détachent de l'image représentée à la figure 2.4. En regardant les images originales, on peut y distinguer le reflet des deux

problèmes que peut rencontrer un système pratique (variations d'éclairage, éclairage de biais, contraste faible ou résolution moins élevée), les améliorations suivantes on été apportées:

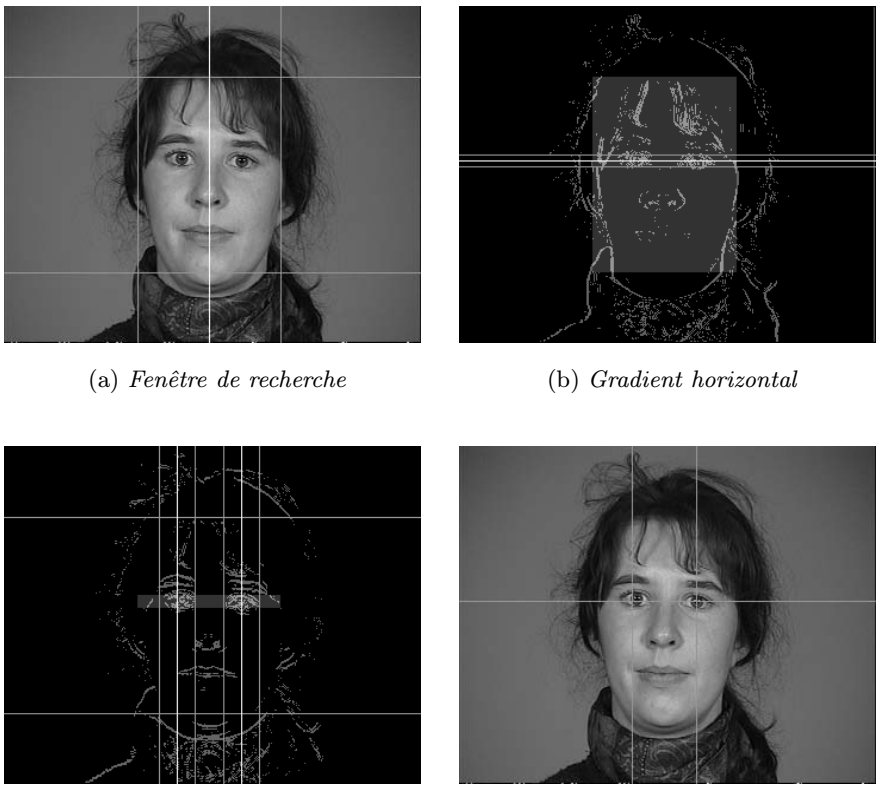
- pour se départir des problèmes liés aux variations d'éclairage, tant globales que locales, les gradients verticaux et horizontaux sont seuillés de façon brutale: un gradient supérieur à un seuil fixé est mis à 1, à 0 dans le cas contraire. Le choix d'un seuil optimal est déterminé par des impératifs contradictoires: il doit être suffisamment bas pour pouvoir détecter un contour dans les zones les plus faiblement contrastées (en général les zones d'ombre), mais suffisamment élevé pour être insensible au bruit de la caméra dans les zones uniformes. En choisissant un bon compromis, on obtient des images gradients binaires robustes aux variations de conditions lumineuses;

- en appliquant la technique de localisation des yeux décrite ci-dessus directement sur les gradients binaires, il est probable que le pic maximum de la PHGH ne corresponde plus à la position verticale des yeux. C'est pourquoi on préférera travailler avec un critère combiné, à savoir détecter, parmi les pics les plus élevés de la PHGH, celui qui donne les pics les plus distincts dans la PVGV;

- enfin, si le fond uniforme est fort contrasté par rapport à la teinte du visage, le contour du visage affectera considérablement la PHGH au détriment des caractéristiques que l'on souhaite isoler. C'est pourquoi, seule la partie centrale du visage doit être prise en considération pour localiser les yeux. La recherche des coordonnées des yeux est alors limitée tant horizontalement que verticalement. Les limites latérales sont liées à la position de l'axe de symétrie du visage (voir figure 2.5(a)). Cet axe, que l'on suppose vertical, est associé à l'abscisse du centre de masse relatif à l'image du gradient horizontal binarisé. Les limites supérieure et inférieure quant à elles, sont fixées une fois pour toutes sur la totalité de la base de données.

La figure 2.5 reprend les différentes améliorations décrites ci-dessus et illustre l'axe de symétrie et les limites de la fenêtre de recherche (a),

spots utilisés pour éclairer le visage. Dès lors, comme les yeux seront caractérisés par des gradients élevés, et ce dans toutes les directions grâce à leur symétrie circulaire, il n'est pas étonnant que la méthode proposée fournisse de très bons résultats.



(a) *Fenêtre de recherche* (b) *Gradient horizontal*

(c) *Gradient vertical* (d) *Localisation des yeux*

Figure 2.5 - *Localisation robuste des yeux*

la localisation du pic maximum de la PHGH sur le gradient horizontal binarisé (b), les deux pics de la PVGV qui y sont associés (c) ainsi que la localisation finale des yeux (d). Lors de la recherche, différents pics de la PHGH sont pris en considération. Ceux-ci sont repris dans le tableau 2.1, généré par le programme de localisation des yeux. Ce tableau fournit les maxima les plus élevés de la PHGH et leurs localisations respectives, la position horizontale éventuelle des yeux (maxima de la PVGV), une mesure de la vraisemblance de cette position[6] et un critère global combinant les valeurs de maximum et de vraisemblance. La localisation finale des yeux est obtenue en sélectionnant la ligne qui offre le meilleur score final (ici la première ligne).

6. Cette mesure est liée à la hauteur et la largeur des deux pics de la PVGV.

```
Displaying the 5 highest peaks :
-------------------------------------------------------
Max = Value         -> (Possible eye      => Final score
         @ line no]     locations) Q=score    (= Max+Q)
-------------------------------------------------------
Max = 30 @ line 131 -> (145,199)   Q=45   => 75 <=
Max = 21 @ line 125 -> (145,205)   Q=24   => 45
Max = 16 @ line  96 -> (138,188)   Q=20   => 36
Max = 16 @ line 137 -> (145,199)   Q=16   => 32
Max = 15 @ line 102 -> (141,194)   Q=12   => 27
-------------------------------------------------------
```

Tableau 2.1 - *Résultats de la recherche du meilleur pic de la PHGH relatif à l'image de la figure 2.5*

La robustesse de la méthode améliorée a ensuite été testée sur une base de données reproduisant des conditions opératoires réelles[7]. La méthode proposée fournit de bons résultats malgré les variations de position, d'échelle, d'éclairage ou d'expression du visage, comme illustrées sur la figure 2.6. La méthode fournit cependant des résultats erronés en présence de lunettes.

7. Il s'agit d'une base de données acquise par MATRA (partenaire industriel du projet M2VTS) dans des conditions opératoires similaires à celles dans lesquelles le produit final est supposé travailler.

Figure 2.6 - *Extraction robuste des coordonnées des yeux dans des conditions opératoires critiques*

Chapitre 3

La base de données M2VTS

3.1 Contexte

Souvent sommes-nous frappés dans la littérature, par le manque d'informations relatives aux conditions sous lesquelles les performances de systèmes d'identification ou d'authentification ont été déterminées: référence succincte de la base de données utilisée, absence de description de la procédure de test suivie, etc. De plus, lorsque information il y a, encore faut-il que celle-ci soit complète: parfois oublie-t-on de préciser des "détails" qui auraient pu se révéler être des conditions nécessaires à l'obtention de bons résultats. Ces propos sont illustrés par quelques exemples. Ainsi [35, 37] omettent de préciser, et ce malgré une description complète de leur base de données (taille, format des images, système d'acquisition), l'intervalle de temps écoulé entre deux prises de vues différentes d'une même personne. On peut se demander si les bonnes performances de leur système ne sont pas liées à un espacement relativement court entre prises de vues. Autre exemple: dans un ensemble d'images, parmi lesquelles se trouvent de nombreuses images prises les unes à la suite des autres, [45, 11] font la distinction entre un ensemble d'entraînement et un ensemble de test, ce qui en soi est un bon point, sans toutefois préciser si les vues successives d'une même personne se retrouvent toutes dans le même sous-ensemble, ou dans les deux sous-ensembles à la fois. Dans ce dernier cas, il n'est pas étonnant que la méthode proposée fournisse d'excellents résultats. Enfin, dans certains cas, toutes les images semblent avoir été acquises au cours de la même session [59, 68, 41].

Dans un souci de rigueur, ce chapitre décrira de façon détaillée la base de données M2VTS utilisée dans ce travail ainsi que le protocole de test suivi pour caractériser les performances de nos différents experts [48, 49].

3.2 Le projet M2VTS

Issu du programme cadre ACTS de la Commission Européenne, le projet européen M2VTS [1] dans lequel s'intègrent les travaux réalisés au cours de cette thèse, traite du contrôle d'accès par authentification biométrique multimodale d'identité. L'aboutissement de ce projet consiste à réaliser un prototype capable d'authentifier l'identité d'une personne sur base de modalités images et paroles novatrices ainsi que d'algorithmes de fusion performants. Deux types de fusion y sont étudiés: la combinaison de résultats provenant de modalités distinctes (comme la voix et le visage vu de face) et, mieux encore, la combinaison directe de différentes modalités au sein d'un même algorithme (l'étude du synchronisme entre la voix et le mouvement des lèvres, par exemple). L'intérêt pour ces nouvelles techniques d'authentification étant relativement récent, aucune base de données de visages existante n'a pu satisfaire les exigences multiples du projet, à savoir offrir les différentes modalités étudiées (tant images que paroles), un son synchrone avec l'image et la possibilité d'extraire des caractéristiques 3-D du visage à partir des images enregistrées. C'est pourquoi le projet a dû procéder à l'enregistrement de sa propre base de données multimodale, la première de ce type à être disponible publiquement.

3.3 La base de données multimodale M2VTS

La base de données M2VTS est constituée de 37 personnes et de 5 prises de vues pour chacune. La cinquième prise de vue reprend une collection de cas plus difficiles à traiter: défauts de mise au point, mauvais rapports signal-à-bruit dans l'image ou le son, yeux clos, visages voilés par une écharpe, présence d'un chapeau, etc. Chaque enregistrement fut espacé d'une semaine au minimum, à moins que des modifications majeures du visage soient intervenues entre-temps. Lors de chaque prise de vue, il a été demandé au sujet

1. *Multi Modal Verification for Teleservices and Security applications.*

de compter de "0" jusqu'à "9" dans sa langue maternelle[2], puis de tourner la tête de façon continue en passant par les positions angulaires suivantes: 0, -90, 0, +90 puis retour à 0 degrés. Si la personne porte des lunettes, il lui est alors demandé de les retirer et de faire ce même mouvement une seconde fois. A partir de cet enregistrement, trois séquences d'images sont extraites: la séquence *voix*, la séquence *mouvement* et la séquence *mouvement sans lunettes* dans le cas où une telle séquence a été filmée. La première séquence peut être utilisée à des fins d'authentification de la parole, de reconnaissance dynamique du visage vu de face (en choisissant automatiquement la ou les images les plus appropriées dans la séquence) ou encore pour l'étude de la corrélation entre la voix et le mouvement des lèvres. Les deux autres séquences sont destinées à des fins d'authentification du visage uniquement et donnent accès à la topologie tridimensionnelle de celui-ci grâce au mouvement de rotation. Ces séquences peuvent être utilisées pour implémenter et comparer des techniques telles que la reconnaissance du visage de face, de profil, de vues intermédiaires ou multiples.

Du matériel offrant une bonne résolution [48] a été utilisé lors de l'enregistrement de la base de données, laissant le choix à l'utilisateur de dégrader la qualité des images par la suite pour simuler un système d'acquisition bon marché. Après diverses conversions de format, la résolution finale des séquences d'images est de 286×350, en 25Hz-progressif. Le son, quant à lui, a été échantillonné à 48 kHz sur 16 bits.

Mis à part le cas de la cinquième prise de vue, cette base de données peut être considérée comme ayant été produite dans des conditions quasi idéales: bonne qualité d'image, enregistrement intérieur, illumination presque constante, fond gris uniforme, etc. Aussi (et surtout) les personnes filmées ont fait de leur mieux pour suivre les instructions qui leur étaient données. Malgré tout, certains écarts par rapport à l'idéal théorique peuvent être remarquées:

- certaines personnes ne parviennent pas à tourner leur tête convenablement et l'on peut noter une translation horizontale du visage dans la direction de la rotation, une inclinaison verticale du visage variable selon l'angle de rotation ou encore une couverture incomplète des 180 degrés;
- certaines personnes peuvent avoir la bouche ouverte lors d'une prise

2. 36 francophones, 1 catalan.

de vue, et fermée dans une autre, débouchant ainsi sur différents contours du profil;

- la direction de départ du mouvement de rotation de la tête peut différer d'une prise de vue à l'autre;
- la focale de la caméra n'a pas été fixée (différents facteurs d'échelle);
- certaines personnes parlent très faiblement, ce qui augmente significativement le niveau de bruit dans la bande son;
- certaines personnes ne peuvent s'empêcher de sourire ou de rire pendant l'enregistrement;
- la vitesse de rotation de la tête peut varier de façon considérable entre les différentes prises de vues, mais aussi au sein de la même séquence;
- un temps d'exposition réduit peut engendrer un flou de "bougé" lors de mouvements de rotation rapides.

Néanmoins, de telles imperfections apparaîtront également en pratique et l'on peut raisonnablement penser qu'un système qui ne parviendrait pas à donner des résultats performants sur cette base de données, ne pourrait pas non plus faire face à des conditions opératoires réelles.

Les figures 3.1, 3.2 et 3.3 illustrent les vues frontales de quelques personnes, la séquence *mouvement* et les 5 prises de vues relatives à un même sujet respectivement.

Les vues de profil et de face utilisées dans les chapitre 1 et 2, ont été sélectionnées manuellement à partir des séquences *mouvement* (voir figure 3.2), avec une tolérance de $0/90 \pm 15$ degrés. Pour information, le profil utile s'étend sur environ 100-150 pixels.

3.4 Protocole de test pour les experts

Jusqu'à présent, tous les tests rapportés dans la littérature et qui font usage de la base de données M2VTS se sont limités aux quatre premières prises de vues [50, 53, 33, 21, 23, 38, 19, 36, 4], laissant de côté la dernière série, plus difficile à traiter, pour des tests ultérieurs. Il en sera de même dans

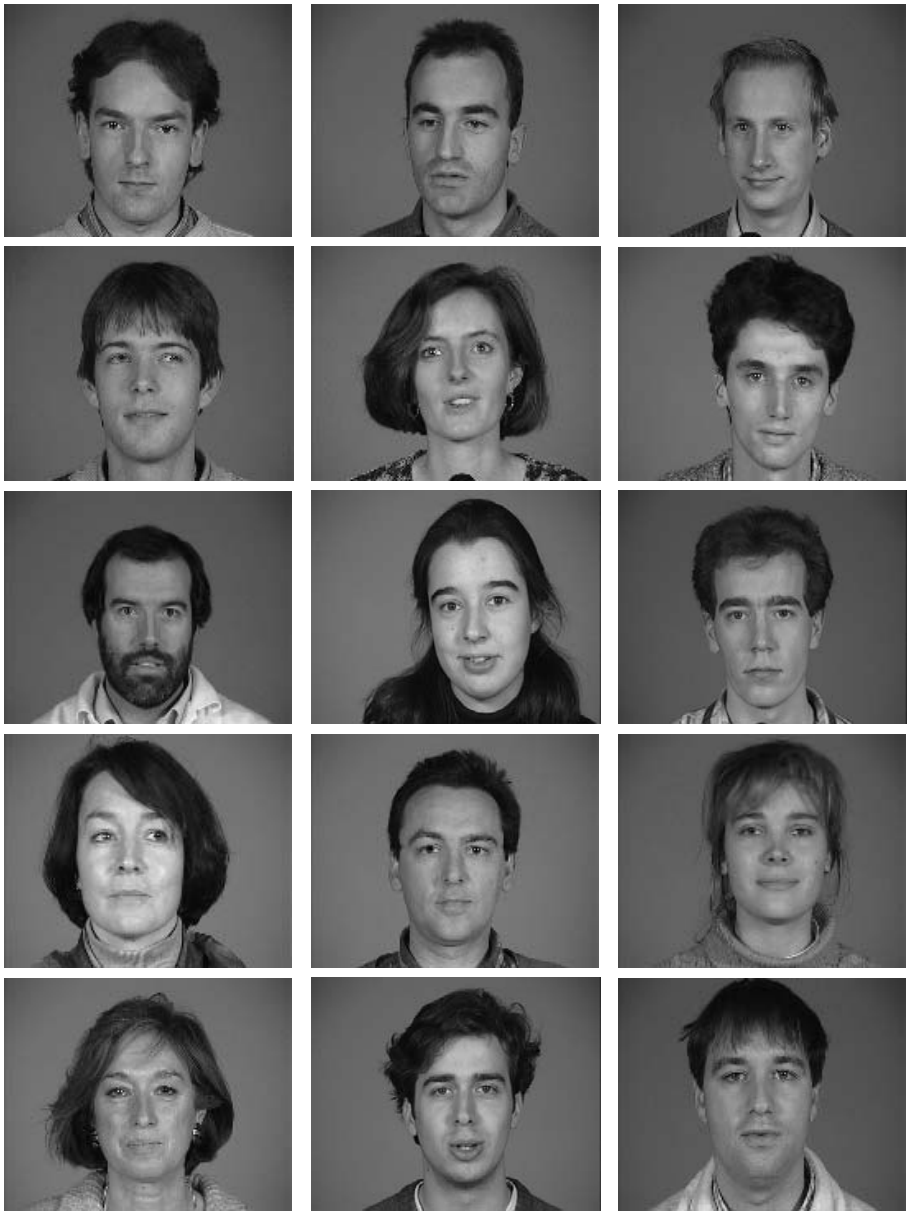


Figure 3.1 - *Base de données M2VTS: quelques vues de face*

Figure 3.2 - *Images extraites d'une séquence "mouvement" (images no. 3, 10, 14, 25, 61 et 75)*

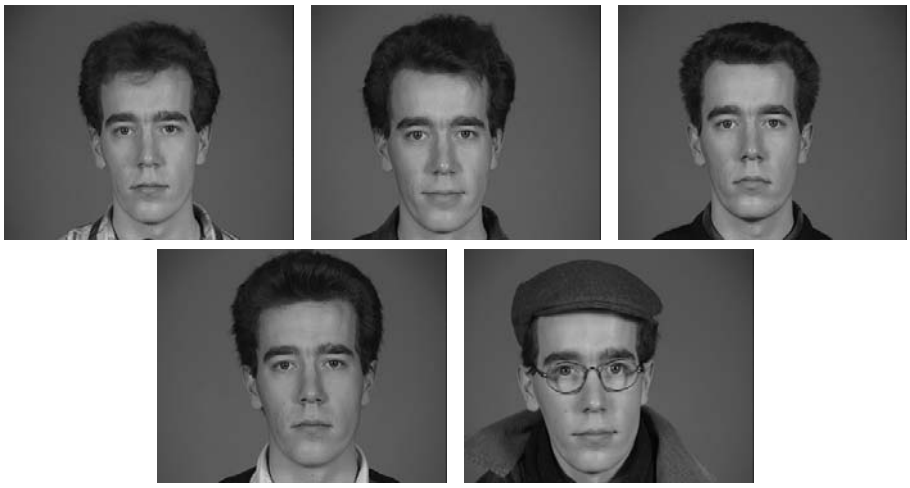


Figure 3.3 - *Les 5 prises de vues relatives à une personne donnée*

le cadre de cette thèse. Tous les tests utilisent un même protocole afin de pouvoir comparer équitablement les performances des différents experts.

Ce protocole définit essentiellement deux ensembles de travail disjoints, à savoir les *ensembles d'apprentissage* et de *test*. L'ensemble d'apprentissage est utilisé pour construire un *modèle* ou sélectionner une ou des *image(s) de référence* pour chaque utilisateur. Le second ensemble sera exclusivement utilisé en phase de test, c'est-à-dire lorsque l'on simule l'accès d'un client ou d'un imposteur en mettant en correspondance l'image de test avec le modèle ou l'image de référence relative à la personne prétendue être.

Dans ses grandes lignes, la procédure suivie pour définir ces deux ensembles se base sur un principe de *mise à l'écart* d'une prise de vue et d'un individu. L'ensemble d'apprentissage est alors constitué de la totalité des données disponibles dans la base de données, à l'exception de la prise de vue et de l'individu mis à l'écart [3]. Ceux-ci fourniront respectivement, lors de l'évaluation des performances d'un expert donné, un ensemble de test client et un imposteur *véritable* [4], comme illustré à la figure 3.4.

En détails, et selon la terminologie adoptée au sein du projet M2VTS [48], la procédure de test complète est décrite comme suit:

> On définit un *ensemble élémentaire de test* comme étant la réunion d'un *ensemble d'apprentissage* [5] et d'un *ensemble de test*. L'*ensemble d'apprentissage* est construit en se choisissant 3 prises de vues (4 sont disponibles) et 36 individus (37 sont disponibles). L'*ensemble de test* est constitué de la prise de vue et de la personne ayant été mises à l'écart lors de la constitution de l'ensemble d'apprentissage. A partir de l'ensemble d'apprentissage, on construit un *modèle de référence* pour chaque client. La performance de l'algorithme d'authentification est alors évaluée en mettant en correspondance chacun des 37 *candidats* issus de l'ensemble de test (36 clients et un imposteur) avec les 36 clients de référence. Un tel ensemble élémentaire de test fournit donc 36 *tests clients* et 36 *tests d'imposture* (l'imposteur contre les 36 clients).

3. et pour rappel, de la cinquième prise de vue dont il ne sera jamais fait usage dans ce travail.

4. Par opposition aux test d'impostures *simulés*, mettant en correspondance deux clients connus du système (appartenant à l'ensemble d'apprentissage).

5. ou d'*entraînement*

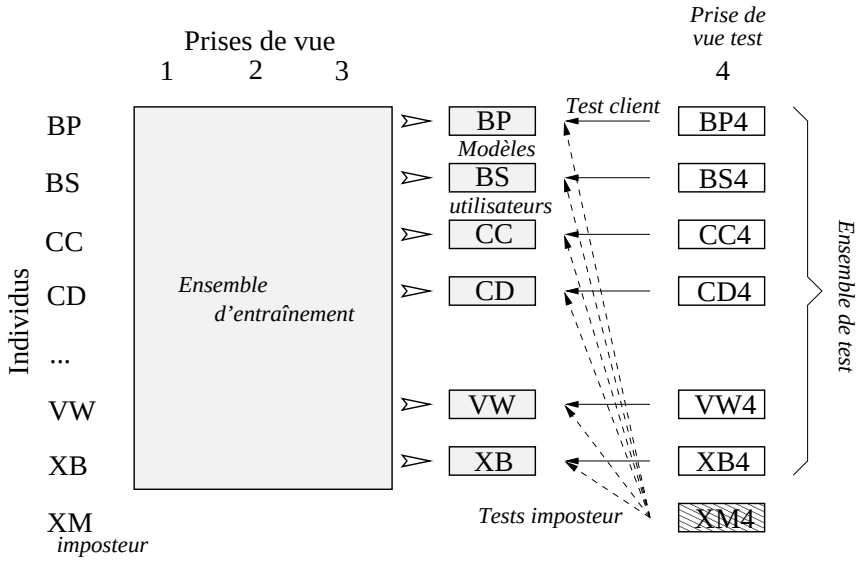


Figure 3.4 - *Procédure de test dans le cas où la personne XM et la prise de vue 4 sont mises à l'écart.*

En effectuant toutes les permutations possibles entre prises de vues et individus, on obtient un total 4×37 ensembles élémentaires de test, soit $4 \times 37 \times 36 = 5328$ tests clients et autant de tests d'imposture. Cet ensemble de 10656 tests est à chaque fois traité lors de l'évaluation des performances de chaque expert.

Hormis l'expert vocal qui sera présenté au chapitre 5, les experts images ne génèrent pas de modèle de référence à proprement parler pendant la phase d'apprentissage, mais font usage de l'ensemble d'apprentissage pour sélectionner une ou plusieurs images représentatives pour chaque client (images de référence). Le candidat issu de l'ensemble de test est alors comparé avec chacune des références de la personne qu'il prétend être. Habituellement, le score obtenu par la meilleure mise en correspondance[6] est utilisé pour authentifier ou rejeter le candidat.

Enfin, signalons que l'ensemble des tests réalisés sur la base de données M2VTS fait usage des images où les utilisateurs ont été priés de *retirer leurs lunettes en face de la caméra*. Ce choix est motivé par deux raisons.

6. soit la plus petite distance résiduelle

La première est liée à la difficulté de pouvoir mettre au point des algorithmes d'extraction de caractéristiques (relief du profil, localisation des yeux) qui soient robustes à la présence éventuelle de lunettes. Ensuite, si le port de lunettes avait pu être traité, une amélioration du taux d'authentification aurait été probable, du moins pour la modalité du visage frontal. En effet, sur une base de données de 37 personnes, il est fort probable que l'on ne retrouve pas deux paires de lunettes identiques. Chaque client pourrait donc être caractérisé par les lunettes qu'il porte. Il suffirait alors que l'algorithme de reconnaisance se focalise plus sur les lunettes que sur le visage proprement dit, pour qu'il en résulte de meilleures performances. Pour pouvoir vérifier la véracité de nos hypothèses, il aurait été nécessaire de pouvoir simuler des accès imposteurs avec des personnes portant le même type de lunettes que le client prétendu être. De telles images n'existent malheureusement pas au sein de la base de données M2VTS, c'est pourquoi nous avons décidé de ne traiter aucune image où des lunettes sont présentes.

Chapitre 4

Performance des différents experts

4.1 Contexte

Ce chapitre caractérise les performances des différentes modalités développées aux chapitres précédents: les deux modalités relatives au profil (sections 1.2.3 et 1.3.4) et la modalité visage de face (chapitre 2). Au cours de ce chapitre, les deux modalités liées à la vue de profil seront combinées et donneront naissance à ce qui par la suite sera appelé l'*expert profil*. Cet expert pourra être a priori considéré comme indépendant de la modalité de face, appelée dorénavant l'*expert frontal*. Cette propriété sera exploitée dans la seconde partie de ce travail relative à la fusion de modalités indépendantes.

4.2 Identification ou Authentification?

Les performances des algorithmes de reconnaissance peuvent être exprimées en termes soit d'identification soit d'authentification.

Au départ, ces deux scénarios abordent des problèmes différents. Dans un problème d'identification, c'est l'identité d'une personne, a priori inconnue, que l'on désire trouver. Tel est le cas lorsque, à partir de la photo d'un agresseur, on tente de retrouver son identité en parcourant les pho-

tos d'une base de données de suspects. Dans un schéma d'authentification, la personne décline son identité. Le système vérifie alors si cette personne correspond bien à celle prétendue être. Un tel schéma est utilisé dans la plupart des applications en surveillance d'accès, que ce soit l'accès à des bâtiments, réseaux informatiques ou services.

Il est néanmoins possible d'authentifier un individu en faisant le détour par un schéma d'identification: lorsqu'un individu se présente devant le système d'authentification, l'identité de la personne de la base de données client qui offre les caractéristiques les plus proches est sélectionnée (identification). Si cette identité correspond à celle prétendue être, la personne est alors authentifiée (authentification). La caractéristique principale d'une telle procédure est de ne recourir à aucun seuil d'acceptation.

Nous avions déjà discuté des avantages et inconvénients respectifs des méthodes d'authentification "proprement dite" ou basées sur une identification préalable (voir l'introduction générale en début d'ouvrage). Nous nous limiterons donc à rappeler ici leurs caractéristiques principales. Celles-ci sont reprises dans le tableau comparatif 4.1. L'avantage majeur offert par un schéma d'authentification proprement dite consiste à offrir des performances indépendantes du nombre de clients traités. C'est pour cette raison, et parce que la taille de la base de données M2VTS est relativement petite, qu'un tel schéma a été adopté dans ce travail. Ce schéma de travail bénéficie aussi d'un temps de calcul fortement réduit par rapport à une procédure d'authentification par identification préalable.

Selon une procédure d'authentification proprement dite, un client verra son accès autorisé si la distance résiduelle résultant de la mise en correspondance de ses caractéristiques biométriques avec celles de la personne qu'il prétend être, est inférieure à un *seuil d'acceptation* k fixé. Dans le cas d'un *seuillage global*, un seuil unique est appliqué sur l'ensemble des clients. La valeur de ce seuil dépend des performances que l'on attend du système: plus k est grand, moins on rejettera de clients, mais plus le nombre d'imposteurs acceptés sera grand. Un compromis doit être choisi en fonction de l'application considérée. Il est également possible de travailler avec un seuil d'acceptation différent pour chaque utilisateur, on parlera alors de *seuillage individuel*. Ces deux techniques seront étudiées dans le courant de ce chapitre.

Schéma	**Authentification par identification**	**Authentification proprement dite**
Performances	Varient selon la taille de la base de données	Indépendantes de la taille de la base de données
Risque d'accepter un imposteur	Peut être plus élevé que dans un système d'authentification mais décroît avec la taille de la base de données	Indépendant du nombre de clients enregistrés. Ne dépend que du seuil d'acceptation que l'on se fixe
Risque de rejeter un client	Peut être plus faible que dans un système d'authentification, mais augmente avec la taille de la base de données	Indépendant de la taille de la base de données. Ne dépend que du seuil d'acceptation que l'on se fixe
Temps calcul	Augmente avec la taille de la base de données (nombreuses comparaisons)	Constant et égal au temps nécessaire pour effectuer une seule comparaison

Tableau 4.1 - *Comparaison entre schémas d'identification et d'authentification*

4.3 Critères de performance

Les critères retenus pour caractériser les performances de nos différentes modalités et experts se basent sur deux mesures fondamentales déjà introduites précédemment, soit le *taux de fausse acceptation* (TFA), c'est-à-dire la proportion d'imposteurs ayant réussi à usurper l'identité d'un client et le *taux de faux rejet* (TFR), la proportion de clients rejetés par le système. Ces mesures sont intimement liées à la valeur du seuil d'acceptation k. Pour rendre cette dépendance plus explicite, la fausse acceptation et le faux rejet peuvent être écrits sous forme de fonctions $FA(k)$ et $FR(k)$. De par leur définition, il apparaît que $FA(k)$ ne peut être qu'une fonction monotone croissante et $FR(k)$ une fonction monotone décroissante[1]. Les critères de

1. Pour rappel, k est un seuil sur une *distance* résiduelle, et non sur un *score* comme défini dans la seconde partie de ce travail.

performance utilisés sont alors les suivants:

- La *courbe caractéristique*[2], qui donne pour chaque valeur de FA, la valeur de FR qui lui est associée. Elle est obtenue en faisant varier continûment le seuil k et en traçant l'ensemble des couples $(FA(k), FR(k))$. Cette courbe fournit de façon graphique un aperçu des tous les compromis TFA/TFR possibles et permet de sélectionner un seuil k adéquat selon l'application envisagée. Une modalité sera d'autant meilleure que sa courbe caractéristique sera proche des axes de coordonnées.
- Le *taux d'égale erreur* (TEE), qui correspond au seuil k' tel que $FA(k') = FR(k') = TEE$. Ce taux, à lui seul, résume assez bien les performances que l'on peut attendre du système.
- Le *taux de succès* (TS), qui équivaut à la valeur maximale de $1 - FA(k) - FR(k)$ et qui fournit une idée des meilleures performances que peut atteindre le système globablement.
- Le taux de faux rejet observé lorsque l'on limite le taux de fausse acceptation à 1% ($\text{TFR}^{1\%}$). Il donne un idée des performances du système lorsque l'application envisagée requiert un taux d'imposture relativement faible.

Caractérisons à présent les performances de nos différentes modalités et experts.

4.4 Seuillage global

Dans cette section, nous ferons part des performances obtenues lorsqu'un même seuil d'acceptation k est appliqué sur l'ensemble des clients. Au départ, nous ferons usage d'une procédure d'entraînement simple qui, à partir des trois images de référence disponibles lors de l'apprentissage (voir section 3.4), sélectionne l'image la plus représentative du client. L'image du candidat est alors comparée à cette unique référence lors de la phase d'authentification proprement dite. Ensuite, nous observerons une amélioration sensible des performances du système si l'image candidate est comparée à

2. *Receiver Operating Characteristics* (ROC) *curve*, en anglais

chacune des trois images disponibles lors de l'apprentissage et que seule la meilleure comparaison est utilisée pour décider du succès ou de l'échec de l'authentification. Naturellement, cette amélioration se fait au détriment d'un temps de calcul trois fois plus important.

4.4.1 Seuillage global avec référence unique (apprentissage)

Pour rappel, dans un tel scénario, un candidat est accepté si la distance résiduelle (distance du chanfrein ou EQM) résultant de la mise en correspondance de l'image candidate avec l'image de référence, est inférieure au seuil d'acceptation k fixé sur l'ensemble de la base de données clients. L'image de référence a préalablement été sélectionnée parmi les trois images disponibles lors de la phase d'apprentissage (voir chapitre 3). Pour ce faire, celles-ci sont comparées les unes aux autres de la façon suivante: $I1$ avec $I2$, $I2$ avec $I3$ et $I3$ avec $I1$, où $I1$, $I2$ et $I3$ désignent les trois images disponibles. En supposant que pour un utilisateur donné, ce soient les mises en correspondance $I1/I2$ et $I3/I1$ qui donnent les meilleurs résultats, l'image $I1$, commune aux deux comparaisons, est alors sélectionnée comme étant l'image la plus représentative du client.

Les courbes caractéristiques relatives à ce test sont illustrées en traits interrompus à la figure 4.1. Aussi, avons-nous introduit la notion d'*expert profil*, qui résulte de la combinaison des deux modalités profils en sommant les scores obtenus par chacune d'elle. Afin de pondérer ces deux modalités de façon équivalente, leur score est préalablement normalisé par le score client moyen calculé sur la base de données d'entraînement. L'expert frontal, quant à lui, est constitué de la seule et unique modalité frontale. Le tableau 4.2 reprend les caractéristiques principales de la figure 4.1.

	TEE	TS	$TFR^{1\%}$
Modalité relief du profil	11.5%	77.5%	59%
Modalité profil niveaux de gris	16.5%	70%	38%
Expert profil	8.5%	83%	30%
Expert frontal	17%	68%	41%

Tableau 4.2 - *Performances des différentes modalités et experts: seuillage global, référence unique.*

On y remarque que les deux modalités basées sur l'analyse des niveaux

de gris offrent des performances similaires et que l'expert profil offre les meilleurs résultats.

4.4.2 Seuillage global avec références multiples

Cette fois, l'image candidate est comparée avec chacune des trois images de référence disponibles. La comparaison qui donne les meilleurs résultats est alors utilisée pour décider du rejet ou de l'acceptation du candidat[3]. Les performances obtenues par un tel schéma sont illustrées à la figure 4.1 et reprises dans le tableau 4.3.

	TEE	TS	$TFR^{1\%}$
Modalité relief du profil	9%	83.5%	50%
Modalité profil niveaux de gris	11%	78.5%	29%
Expert profil	8%	85.5%	18.5%
Expert frontal	9%	83%	23%

Tableau 4.3 - *Performances des différentes modalités et experts: seuillage global, références multiples.*

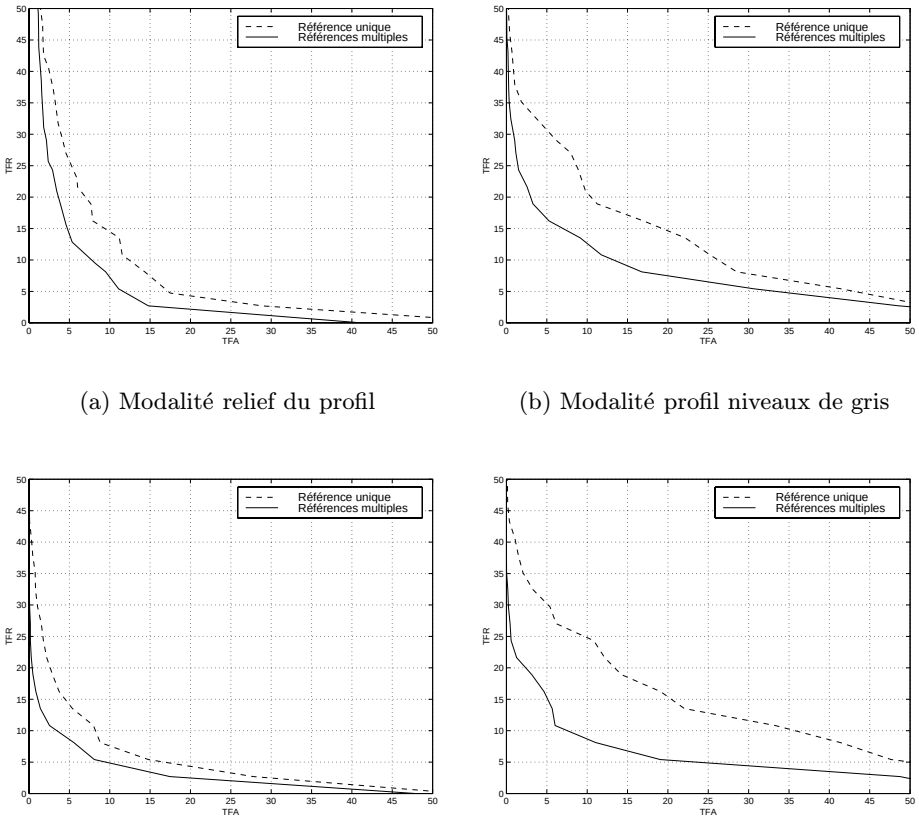
L'utilisation de références *multiples* permet de réduire de près de la moitié le taux de faux rejet pour un taux de fausse acceptation donné. Cette amélioration se fait au détriment d'un temps de calcul trois fois plus important: d'environ 0.5 seconde il passe à 1.5 seconde pour chacun des deux experts[4]. Ce temps restant toutefois raisonnable, il sera toujours fait usage d'une telle référence dans la suite de ce travail.

4.4.3 Distinction entre profils entiers et partiels

Les performances données jusqu'à présent font usage d'un seuil d'acceptation k unique appliqué à l'ensemble des clients, ce qui revient encore à traiter tous les clients sur un même pied d'égalité. En pratique, néanmoins,

3. Des tests faisant usage de la moins bonne comparaison ont été effectués, mais se sont avérés catastrophiques. Il est en effet possible pour un imposteur d'offrir une "plus mauvaise mise en correspondance" qui soit meilleure que la plus mauvaise mise en correspondance du client. Ce n'est pas le cas en prenant la meilleure comparaison comme base de décision.

4. Tests effectués sur un PC Pentium 100 Mz.



(a) Modalité relief du profil

(b) Modalité profil niveaux de gris

(c) Expert profil

(d) Expert frontal

Figure 4.1 - *Performances des différentes modalités et experts dans le cas d'un seuillage global: référence unique (traits discontinus) et références multiples (traits continus)*

certains clients pourront plus facilement être sujets à une imposture que d'autres, selon les caractéristiques biométriques qui les caractérisent. Pour éviter ceci, il faudra choisir un k suffisamment faible, pénalisant ainsi tous les autres clients[5]. Ce problème apparaît clairement pour les modalités liées au profil. En effet, si l'extraction complète du contour du profil n'est pas possible, seule une partie de celui-ci est effectivement utilisée (voir les différents modes d'extraction, section 1.4.2). Les profils dits *partiels* (Modes 2, 3 et 4) souffriront d'un TFA plus élevé puisque il est plus facile de mettre en correspondance des zones restreintes du profil, même ci celles-ci appartiennent à deux personnes différentes, que des profils entiers. Mieux vaut donc pouvoir traiter ces profils séparément et leur associer un seuil k' propre[6]. La figure 4.2 et le tableau 4.4 illustrent ce propos et donnent les performances des différents algorithmes sur les deux sous-ensembles *profils partiels* (Modes 2, 3 et 4) et *profils complets* (Mode 1) séparément. Bien que cette distinction se base sur la vue de profil, l'expert frontal semble lui aussi y être sensible et éprouve plus de difficultés sur le sous-ensemble relatif aux profils partiels[7]. Ceci s'explique par la présence de visages dont les cheveux tombent au niveau des yeux (engendrant un profil partiel de mode 2) et qui de ce fait empêchent une bonne mise en correspondance des fenêtres actives frontales.

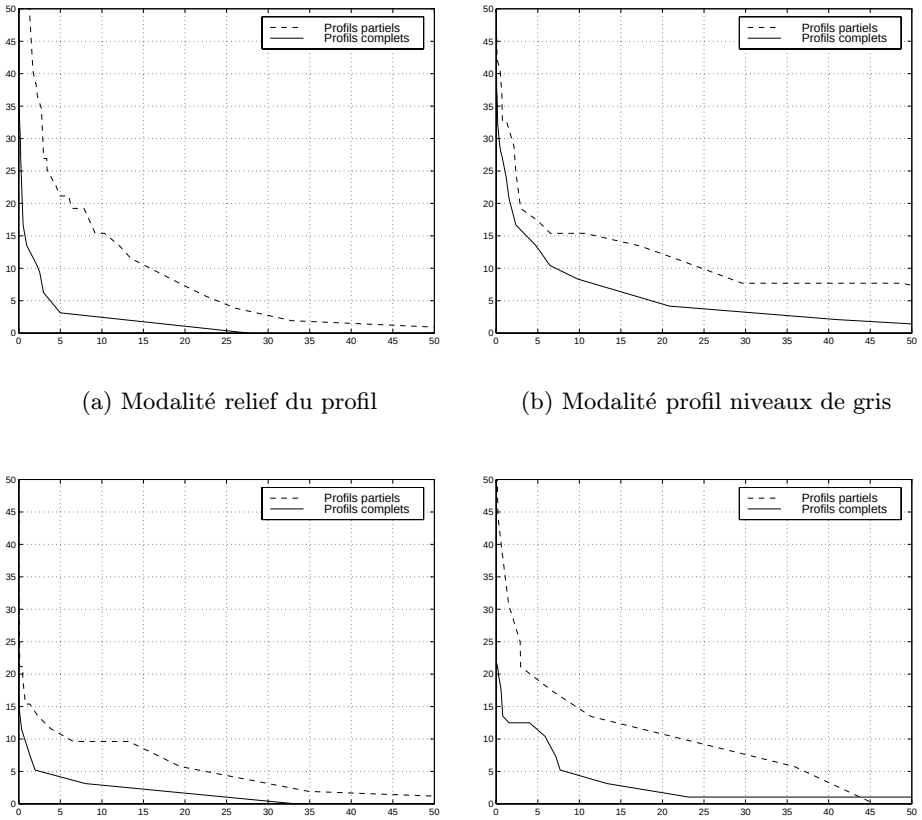
		TEE	TS	$TFR^{1\%}$
Modalité relief du profil	PP	12.5%	76%	54%
	PC	4%	92%	13.5%
Modalité profil niveaux de gris	PP	14%	78%	32.5%
	PC	9%	81%	25%
Expert profil	PP	9.5%	84.5%	15.5%
	PC	4.5%	93%	9%
Expert frontal	PP	13%	76%	36%
	PC	7%	87%	13%

Tableau 4.4 - *Performances des sous-ensembles "profils partiels" (PP) et "profils complets" (PC). [Références multiples]*

5. En effet, pour les clients dont les caractéristiques varient relativement beaucoup, il aurait fallu utiliser un seuil élevé pour éviter qu'ils soient trop souvent rejetés.

6. Pour information, on compte sur les 37 personnes de la base de données, 13 personnes caractérisées par un profil partiel pour 24 profils complets.

7. Les deux experts ne seraient donc pas aussi indépendants que ce que l'on aurait pu croire (cfr. début du chapitre).



(a) Modalité relief du profil

(b) Modalité profil niveaux de gris

(c) Expert profil

(d) Expert frontal

Figure 4.2 - *Comparaison des performances atteintes au sein des sous-ensembles "profils complets" (traits continus) et "profils partiels" (traits discontinus) [Références multiples]*

4.5 Seuillage individuel

Nous avons pu voir, dans la section précédente, l'avantage qui résulte de la prise en compte des disparités existantes entre différents groupes d'utilisateurs. Mieux encore, nous pouvons offrir à chaque client un seuil d'acceptation propre qui tient compte de la facilité ou de la difficulté qu'un imposteur aurait à usurper son identité.

Ces seuils individuels sont obtenus durant la phase d'entraînement de chaque algorithme (c'est à dire en faisant usage de la base de données d'apprentissage exclusivement: 36 clients/3 prises de vues) de la façon suivante:

- pour chaque client, on considère les 35 autres comme étant des imposteurs et on fixe le seuil d'acceptation pour ce client comme égal à la moyenne des scores relatifs aux 5 meilleurs accès imposteurs;
- les scores imposteurs sont calculés en prenant le meilleur score issu de toutes les mises en correspondance possibles entre les 3 références imposteur avec les 3 références client.

Plutôt que de prendre en considération les meilleurs accès imposteurs, nous aurions pu travailler avec les plus mauvais accès clients. Néanmoins, caractériser le comportement des imposteurs est plus robuste que celui des clients, la quantité de données relatives aux tests d'imposture étant de loin plus étendue. Le calcul de la moyenne des cinq meilleurs scores imposteurs permet quant à lui de mieux caractériser le comportement des imposteurs les plus dangereux et offre de meilleurs seuils que ceux obtenus par une moyenne sur l'ensemble des imposteurs, ou en tenant uniquement compte du meilleur imposteur.

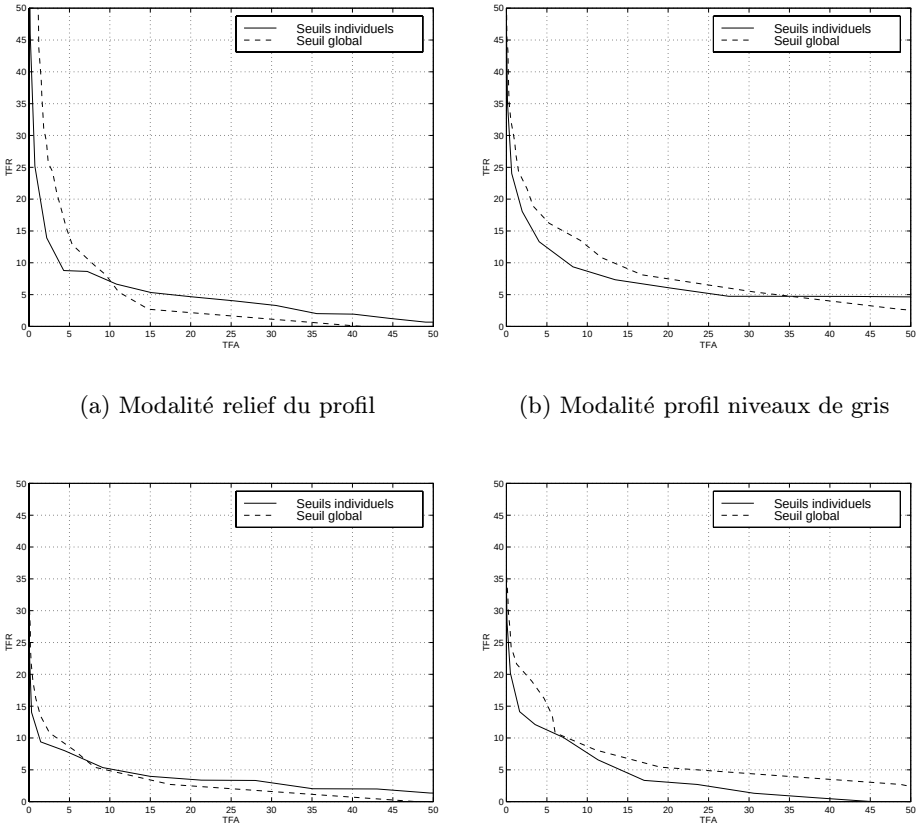
La procédure qui vient d'être décrite fournit 36 seuils individuels et un seul point $(TFA; TFR)$ dans la courbe caractéristique de chaque modalité/expert. Afin de pouvoir accéder à d'autres conditions opératoires et pouvoir tracer une courbe caractéristique complète, d'autres vecteurs de seuils individuels ont été obtenus en appliquant différents facteurs d'échelle au vecteur initial (facteurs allant de 0.5 à 5). Les courbes caractéristiques de chaque modalité ainsi obtenues sont reprises à la figure 4.3. Le tableau 4.5, résume quant à lui les principales performances.

	TEE	TS	$TFR^{1\%}$
Modalité relief du profil	8%	87%	23%
Modalité profil niveaux de gris	9%	83%	22.5%
Expert profil	7%	89%	11%
Expert frontal	8.5%	84%	17.5%

Tableau 4.5 - *Performances des différentes modalités et experts : seuillage individuel, références multiples*

Sans véritable surprise, l'utilisation de seuils individuels permet d'obtenir de bien meilleurs résultats, puisqu'elle permet de relâcher la contrainte, c'est-à-dire relever le seuil d'acceptation, pour des clients plus difficiles à imposter. A taux de fausse acceptation fixé, l'utilisation de seuils individuels permet de réduire le taux de faux rejet de 30% environ.

Enfin, remarquons la dissymétrie qui existe entre le taux de faux rejet observé pour un taux de fausse acceptation nul et le taux de fausse acceptation observé à faux rejet nul, ce dernier étant de loin plus élevé (figure 4.3). On pourrait croire à première vue qu'en fixant les seuils individuels par rapport aux imposteurs, on apprend à mieux rejeter ces derniers qu'à accepter les clients. En fait, ce déséquilibre est inhérent au problème même de l'authentification d'identité. Il est en effet difficile pour un imposteur de ressembler fidèlement au client qu'il imposte, mais facile pour un client d'arriver à "ne pas se ressembler", c'est-à-dire à s'écarter fortement de son modèle: il lui suffit par exemple de mal se positionner devant le système. Si l'on veut néanmoins que le système puisse accepter de tels clients (tel est le cas lorsqu'on travaille à taux de faux rejet nul), nous devrons placer le seuil d'acceptation relativement haut et accepter de ce fait un nombre d'imposteurs relativement élevé.



(a) Modalité relief du profil

(b) Modalité profil niveaux de gris

(c) Expert profil

(d) Expert frontal

Figure 4.3 - *Performances des différentes modalités et experts dans le cas d'un seuillage individuel sur l'ensemble de la base de données [Références multiples]*

Chapitre 5

Un expert supplémentaire : la voix

5.1 Contexte

Ce chapitre introduit un nouvel expert basé sur une modalité différente des modalités images décrites jusqu'à présent: la parole. Développé par l'IDIAP [1], partenaire de l'UCL au sein du projet M2VTS, cet expert s'écarte quelque peu des travaux spécifiquement réalisés dans le cadre de cette thèse. Il a néanmoins été jugé utile d'y faire référence pour les excellentes performances offertes et la technique prometteuse sur laquelle il se base. Aussi, s'intégrera-t-il parfaitement dans le superviseur développé dans la seconde partie de ce travail, fournissant une modalité supplémentaire indépendante de celles développées au sein de l'UCL.

1. *Institut Dalle Molle d'Intelligence Artificielle Perceptive*, Martigny, Suisse

5.2 Architecture générale

L'expert vocal développé par l'IDIAP se base sur les modèles de Markov[2] cachés [33], méthode actuellement reconnue comme performante et dont l'utilisation s'est largement répandue en authentification automatique de la voix [26]. A l'instar de techniques dites *indépendantes du texte*, le mode de vérification utilisé demande au client de prononcer une phrase connue du système, en l'occurrence la série de chiffres de 0 jusqu'à 9, telle que prononcée dans la base de données M2VTS. Une telle technique est qualifiée de *dépendante du texte*.

Lors de la phase d'authentification, deux modèles distincts sont utilisés pour chaque chiffre: un *modèle universel*[3] acquis sur un grand nombre d'utilisateurs différents et qui représente en quelque sorte le monde extérieur et un *modèle client* propre à l'utilisateur considéré. Pour chaque "chiffre" extrait de la séquence de parole candidate, les vraisemblances relatives aux modèles universel et client sont mesurées. Les vraisemblances obtenues sur chaque chiffre sont alors multipliées entre elles pour obtenir une mesure globale. Si cette mesure est sensiblement plus élevée pour le modèle client que pour le modèle universel, l'utilisateur candidat est authentifié. Les sections suivantes expliquent comment de tels modèles peuvent être obtenus.

5.3 Les modèles de Markov cachés

De par leur définition, les phonèmes peuvent être vus comme les éléments de base qui forment la parole. Tous différents les uns des autres, ils peuvent néanmoins varier selon les sons qui les précèdent ou les suivent. Ainsi est-il commode de définir trois états au sein de chaque phonème: un état initial,

2. Andrei Andreyevich Markov, né en 1856 à Ryazan (Russie), mort en 1922 à Saint-Petersbourg (Russie), fut diplômé de l'université de cette même ville en 1878 et y devint professeur en 1886. Ses premiers travaux furent principalement dédiés à la théorie des nombres, limites d'intégrales, théorie des approximations et convergence des séries mathématiques. Suivant les traces de son professeur Pafnuty Chebyshev, il s'attaqua à la théorie probabiliste vers 1900 et en particulier à l'étude des variables mutuellement dépendantes. Son nom est associé aux chaînes de Markov, séquences de variables aléatoires dans lesquelles la valeur future d'une variable est déterminée par sa valeur actuelle uniquement. Markov est considéré comme l'un des principaux précurseurs de la théorie des processus stochastiques.

3. *World model*, en anglais.

qui tient compte de la transition entre le son précédent et le début du phonème, un état mitoyen et un état final, amorçant le début du phonème suivant.

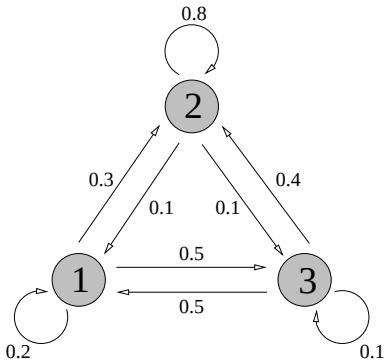


Figure 5.1 - *Exemple d'une chaîne de Markov (3 états)*

Une chaîne de Markov est constituée d'un certain nombre de nœuds, liés entre eux par des transitions possibles [42]. A chaque transition est associée une probabilité et à chaque nœud, une valeur de la variable que l'on souhaite modéliser. En toute généralité, cette variable peut être multidimensionnelle. Une telle chaîne est illustrée à la figure 5.1.

La figure 5.2 illustre l'utilisation la modélisation du phonème [è] (du mot *thèse*) et les trois états mentionnés ci-dessus. Seules quelques transitions sont autorisées, soit les transitions d'un état vers lui même (permettant d'introduire une dilatation temporelle et de tenir compte de la vitesse avec laquelle le son est prononcé) et celles d'un état vers un état futur (il n'est en effet pas possible de remonter dans le temps).

Une modélisation encore meilleure peut être obtenue par l'usage d'un *Modèle de Markov Caché* (MMC)[4]. Un MMC est semblable à une chaîne de Markov, si ce n'est qu'à chaque nœud sont associées toutes les valeurs possibles de la variable à modéliser, valeurs dont les occurrences sont régies par une fonction de densité de probabilité caractéristique au nœud. Un exemple de MMC est repris à la figure 5.3. Cette figure représente un phonème décomposé en intervalles de temps réguliers (typiquement 25ms). Le spectre

4. *Hidden Markov Models* (HMM), en anglais.

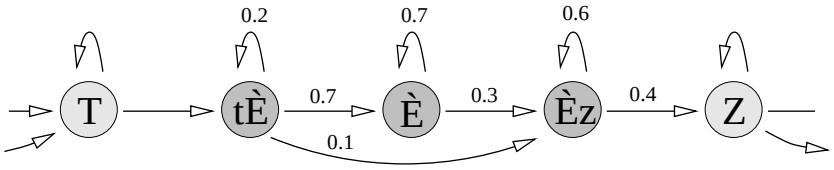


Figure 5.2 - *Modélisation d'un phonème par une chaine de Markov à 3 états*

de la parole est calculé en chacun de ces intervalles et diverses caractéristiques de la voix en sont extraites. Ce sont ces caractéristiques et leur évolution au cours du temps qui, par l'intermédiaire des différentes densités de probabilité associées à chaque nœud et des probabilités de transition entre nœuds, sont modélisées par le MMC. Parmi les caractéristiques les plus communément extraites, nous retrouvons les coefficients cepstraux et leur prédiction linéaire[5] [25, 26].

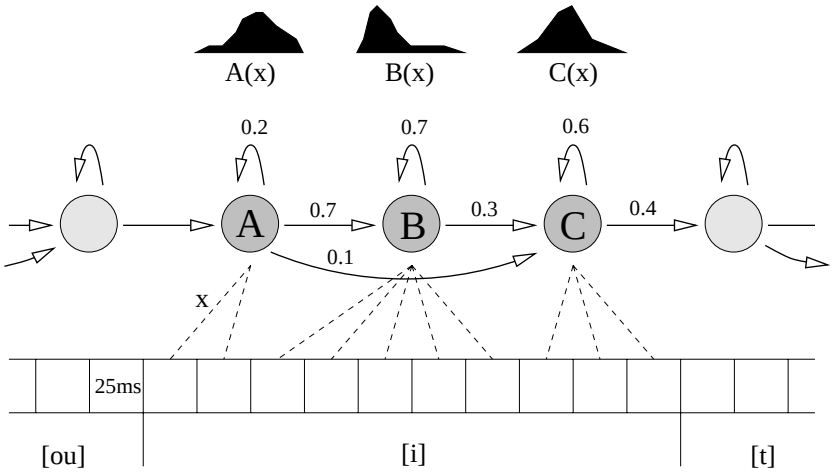


Figure 5.3 - *Modèle de Markov caché, appliqué à la reconnaissance de phonèmes*

Chaque phrase peut être décomposée en phonèmes et faire l'objet d'une modélisation par MMC. Si l'on entraîne le MMC sur suffisamment de per-

5. *Linear Prediction Cepstral* (LPC) *coefficients*, en anglais

sonnes différentes, on obtiendra une représentation relativement indépendante du locuteur. Ainsi, pour modéliser la succession des chiffres "0", "1", ..., "9" utilisée lors de l'authentification, une base de données de plus de 500 personnes fut utilisée par l'IDIAP (base de donnée *Polyphone*, voir [15]). Une fois ce modèle acquis, il est possible de segmenter automatiquement un signal de parole donné en ses diverses composantes, soit les différents chiffres dans le cas qui nous préoccupe.

Pour ce faire, le signal de parole tout entier est segmenté en intervalles de 25ms. Chaque intervalle est alors rattaché à un état particulier d'un phonème composant l'un des dix chiffres. La figure 5.3 illustre le phonème [i] du chiffre [ouit]. On peut calculer, pour ce phonème particulier, la probabilité qu'il ait effectivement été décomposé comme suggéré dans la figure, c'est-à-dire en associant les deux premiers tronçons au premier état, les 5 suivants à l'état intermédiaire et les 3 derniers à l'état final. Cette probabilité équivaut simplement au produit des différentes probabilités de transition entre nœuds combinées aux probabilités d'observation des différents coefficients cepstraux, liées aux densités de probabilité associées à chaque nœud. Le recalage exact du phonème sera obtenu en maximisant sa probabilité d'observation. Pour bien faire, il faut alors envisager toutes les mises en correspondance possibles entre intervalles de temps et nœuds du MMC, calculer leurs probabilités respectives puis sélectionner la mise en correspondance la plus probable. Cette tâche peut rapidement devenir fastidieuse, comme le suggère la figure 5.4 qui illustre l'ensemble des trajets possibles pour relier un nœud à l'autre ou à lui-même.

Heureusement, il y a moyen de tenir compte de la structure particulière des MMC (à savoir, que la probabilité d'entrer dans un état dépend uniquement de l'état précédent), et de simplifier le schéma de la figure 5.4 par le *treillis* présenté à la figure 5.5. En appliquant un algorithme tel celui de Viterbi [30], il n'est plus nécessaire d'envisager toutes les combinaisons possibles. Cet algorithme travaille de façon itérative: à chaque étape, un intervalle de temps supplémentaire est pris en compte et l'on regarde si celui-ci est source de "collisions" dans le treillis. Si tel est le cas, la probabilité de tous les chemins qui mènent à de telles collisions est alors calculée. Le chemin le moins probable est alors supprimé définitivement du treillis, comme illustré à la figure 5.6. Cette technique permet de ne retenir que les chemins les plus probables à chaque étape. En fin de parcours, le meilleur de ceux-ci est alors sélectionné.

Le recalage de la phrase entière peut être effectué de la même façon, ce

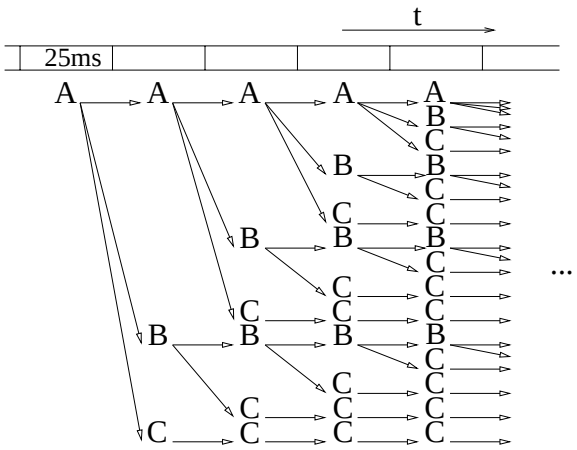


Figure 5.4 - *Diagramme de toutes les assignations possibles au sein d'un même phonème*

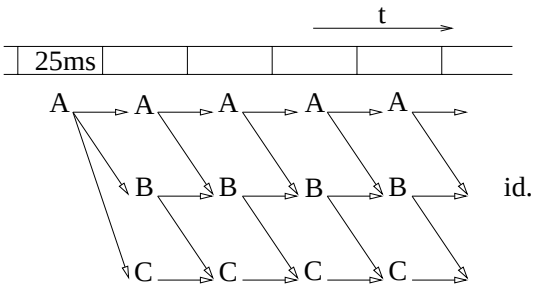


Figure 5.5 - *Treillis associé à la figure 5.4*

qui permet de localiser précisément le début et la fin de chaque chiffre. Une fois cette segmentation réalisée, il ne reste plus qu'à travailler sur chaque chiffre pris séparément. Ayant préalablement entraîné un MMC spécifique à la voix du client on compare alors les probabilités que chaque chiffre extrait de l'échantillon de parole candidat appartienne soit au MMC candidat, soit au MMC universel (celui utilisé lors de la segmentation). Si la première probabilité est suffisamment plus élevée que la seconde, le client est authentifié.

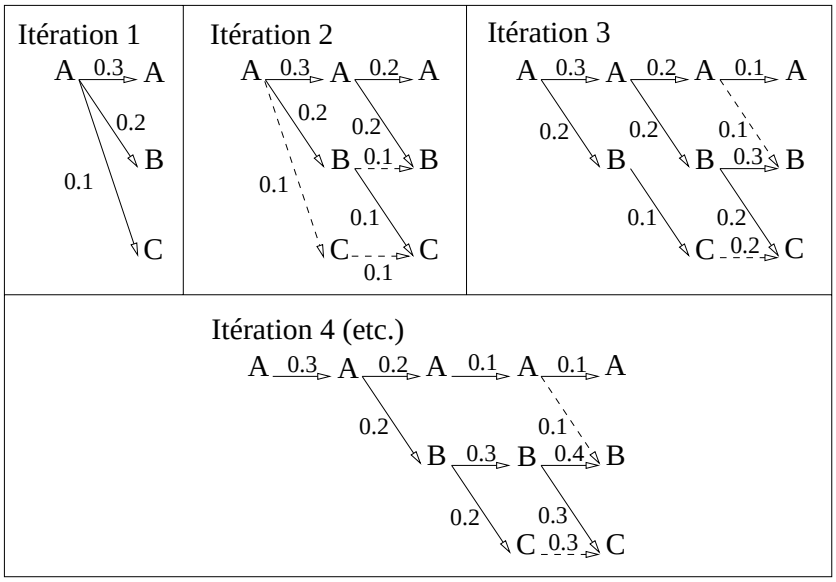


Figure 5.6 - *Algorithme de Viterbi (en traits interrompus, les chemins définitivement supprimés à chaque itération)*

5.4 Acquisition du modèle client

L'algorithme de Viterbi peut également être utilisé lors de l'acquisition du modèle client (à savoir les diverses densités de probabilité associées en chaque nœud du MMC ainsi que les probabilités de transition entre nœuds). Au départ, ce algorithme sert à recaler des données d'entraînement propres au client, sur un modèle initial. Une fois recalées, ces données permettent d'affiner le modèle, qui servira à recaler avec plus de précision de nouvelles données d'entraînement, et ainsi de suite. Ce processus itératif s'arrête lorsque un modèle client est obtenu avec suffisament de précision. De plus amples informations relatives à l'entraînement de modèles de Markov cachés peuvent être trouvées dans [17].

5.5 Performance de l'expert vocal

Le modèle universel ayant été entraîné sur une base de donnés acquise à travers un réseau téléphonique, le signal de parole issu de la base de données M2VTS est d'abord dégradé pour offrir la même qualité sonore (8 kHz/8 bits). Malgré cela, l'expert vocal offre des performances tout à fait remarquables comme le montre le tableau 5.1.

	TEE	TS	$TFR^{1\%}$
Expert parole IDIAP	1.5%	97.5%	1.5%

Tableau 5.1 - *Performances de l'expert vocal développé par l'IDIAP [seuillage individuel]*

L'algorithme travaille sur des intervalles de 25 ms, espacés de 10 ms chacun. Sur chaque intervalle, 39 coefficients cepstraux sont isolés et permettent de caractériser les différents états des MMC. Les MMC utilisés font usage de 2 à 7 états selon le chiffre modélisé [33]. La courbe caractéristique de l'expert vocal est illustrée à la figure 5.7[6].

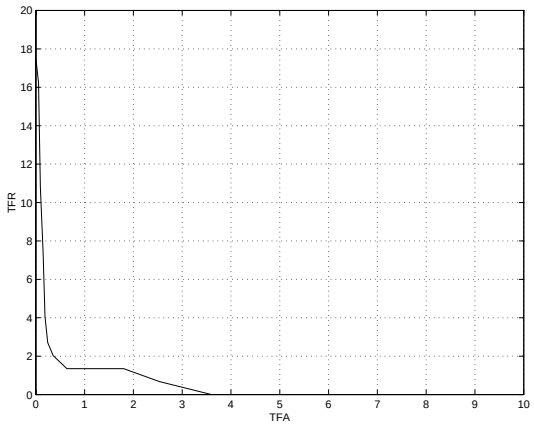


Figure 5.7 - *Courbe caractéristique de l'expert vocal développé par l'IDIAP [seuillage individuel]*

6. Notez le changement d'échelle par rapport aux figures similaires du chapitre 4

Chapitre 6

Conclusion de la première partie

Cette première partie fut consacrée à l'élaboration de méthodes d'authentification d'identité basées sur différentes caractéristiques biométriques liées au visage et à la voix, ainsi qu'à la caractérisation de leurs performances. Une attention toute particulière fut portée à l'usage d'algorithmes rapides caractérisés par un excellent compromis entre la simplicité de leur mise en œuvre et le niveau de performance offert. L'algorithme du simplexe fut combiné à celui du chanfrein pour déboucher sur une méthode de mise en correspondance qui, après diverses optimisations, fut utilisée avec succès au sein des experts profil et frontal. L'authentification basée sur le relief du profil s'est vue adjoindre un module travaillant sur les niveaux de gris relatifs à la même vue. Ceci eut pour effet d'améliorer sensiblement les performances de l'expert profil sans toutefois en augmenter la complexité algorithmique, le module en niveaux de gris faisant usage des mêmes paramètres de compensation que ceux utilisés pour la mise en correspondance du contour du profil.

Nous avons également été amenés à résoudre des problèmes plus pratiques comme l'extraction de contours sur des images couleur à faible contraste (images de profil) ou la localisation automatique d'une fenêtre centrée autour des yeux (images de face). Ces problèmes résolus, les méthodes proposées ne nécessitaient plus aucune intervention manuelle autre que la sélection des images profils ou frontales représentatives du client. Dès lors, les performances mentionnées dans le présent travail sont représentatives de

l'ensemble de la chaîne de traitement automatisée qui comprend à la fois les algorithmes d'authentification et le pré-traitement des images d'entrée.

Afin d'évaluer ces performances aussi rigoureusement que possible, un protocole de test fut défini sur une base de données proche de celle que nous pourrions être amenés à traiter en réalité. Ce protocole définissait des ensembles d'apprentissage et de tests distincts ainsi que des imposteurs véritablement inconnus de l'ensemble d'entraînement. Par un jeu de rotations entre individus d'une part et entre prises de vues d'autre part, plus de 10,000 tests clients et imposteurs furent réalisés pour caractériser les performances d'un algorithme particulier. Ces performances ont été évaluées selon trois scénarios distincts. Dans le premier, chaque client avait été représenté par une prise de vue unique, sélectionnée de façon automatique parmi les prises de vues disponibles durant la phase d'apprentissage. Dans le second, l'ensemble de ces prises de vues fut mis à disposition des divers modules d'authentification. Ceci avait permis d'accroître considérablement les performances du système, au détriment toutefois d'un temps de calcul plus élevé. Enfin, le troisième scénario ajoutait au second l'usage de seuils d'acceptation individuels. Il offrait de loin les meilleures performances, reprises dans le tableau ci-dessous:

	TEE	TS	$TFR^{1\%}$
Modalité relief du profil	8%	87%	23%
Modalité profil niveaux de gris	9%	83%	22.5%
Expert profil	7%	89%	11%
Expert frontal	8.5%	84%	17.5%
Expert vocal	1.5%	97.5%	1.5%

Tableau 6.1 - *Récapitulatif des performances des différentes modalités et experts dans le cas d'un seuillage individuel et d'une référence multiple*

Enfin, cette première partie se termina sur la description d'un algorithme d'authentification de la voix développé par l'IDIAP. Cet expert, jouissant de propriétés remarquables, ne pouvait être omis dans ce travail. Il viendra compléter les experts images profil et frontal au sein du superviseur développé dans la seconde partie de ce travail.

Deuxième partie

Le Superviseur

Chapitre 1

Approche mathématique

1.1 Contexte

Ce premier chapitre formalise le problème de l'authentification d'identité au sein du contexte de la fusion de données. Après avoir présenté les deux grandes familles de superviseurs étudiés par la suite (section 1.2), nous formulerons mathématiquement le problème posé (section 1.3). Pour le résoudre, différentes approches seront envisagées, à savoir les approches de Neyman-Pearson (section 1.4) et de Bayes (section 1.5). Toutes deux déboucheront sur une même forme de solution optimale. Enfin, le formalisme développé dans ce chapitre nous permettra de détailler la spécificité de chaque superviseur étudié dans le cadre de cette seconde partie (sections 1.6 et 1.7).

1.2 Les deux grandes classes de fusion

On distingue généralement deux grandes familles de superviseurs[1] selon la nature de l'information que transmettent les différents experts. Toutes deux seront traitées dans le cadre de ce travail.

Dans la première famille, chaque expert remet son avis quant à l'accepta-

1. Pour rappel, le superviseur est le module chargé d'implémenter la fusion des données en provenance des différents experts.

tion ou le rejet de l'individu qui demande l'accès au système. Ces données binaires (acceptation/rejet) sont traitées par le superviseur qui, ayant une vue d'ensemble des différentes opinions, est à même de prendre au mieux la décision finale. Puisque les données à traiter sont binaires, la fusion s'effectue généralement par l'intermédiaire d'opérateurs logiques (opérateurs ET ou OU par exemple) [2]. Si le nombre d'experts le permet, d'autres règles de décision peuvent être utilisées comme le choix de l'avis majoritaire ou de l'avis le plus vraisemblable (choix plus nuancé qui résulte du calcul d'un rapport de vraisemblance – voir section 1.6). De tels schémas de fusion sont appelés *fusions de décisions*, faisant référence aux décisions que prennent les différents experts que l'on fusionne, ou encore *fusion dure*, en regard des données "brutes" binaires traitées. Le terme *fusion aval* est parfois utilisé dans la littérature et signifie que l'on fusionne *après* que les experts aient remis leur avis.

La deuxième famille de superviseurs fait usage direct des *scores* fournis par chacun des experts et les combine en un score global qui, seuillé, fournira la décision finale d'accepter ou de rejeter le candidat. Cette combinaison peut se faire selon n'importe quelle fonction des scores issus des experts: depuis une simple combinaison linéaire jusqu'aux règles moins méthodiques de la logique floue [31], en passant par des règles de décision optimisant un critère statistique particulier, tel le critère de Bayes étudié par la suite. On peut également ranger dans cette catégorie des classificateurs faisant explicitement usage de l'ensemble des données d'apprentissage comme le classificateur k-NN[2] par exemple [64]. Pour désigner l'ensemble des superviseurs appartenant à cette seconde famille, les termes *fusion de scores*, ou *fusion douce* sont communément utilisés. D'autres préfèrent le vocable *fusion en amont*, faisant référence au fait qu'aucune décision n'a été prise avant d'effectuer la fusion proprement dite [2].

Un schéma de fusion basé sur une combinaison de scores (*fusion douce*) est capable a priori d'offrir de meilleures performances qu'une fusion basée sur le regroupement de décisions individuelles (*fusion dure*). En effet, on peut considérer qu'à chaque prise de décision – ce qui revient à seuiller un score ou une distance résiduelle – une partie de l'information disponible sur la véracité de l'identité d'un individu est perdue. Ainsi, un score chiffré tel 0.567 contient plus d'information qu'un résultat binaire tel que *1=accepté* et *0=rejeté*. Le score 0.567 indique que l'on est en présence d'un cas critique

2. *k-Nearest Neighbours*, soit les k plus proches voisins

et que si l'on accepte le candidat, cela se fait de justesse[3]. On peut alors prendre des précautions supplémentaires, par exemple demander l'avis d'experts additionnels. Il vaut donc mieux travailler autant que possible dans le domaine des scores et n'effectuer un seuillage que lors de la décision ultime. Ceci motive l'utilisation d'une fusion douce. Néanmoins, et pour autant que l'on se limite à des techniques simples de fusion douce (comme la combinaison linéaire des scores, étudiée dans le cadre de ce travail), une fusion dure peut offrir de meilleures performances selon la nature des données que l'on fusionne et le point opératoire que l'on désire atteindre. Ce phénomène sera illustré par la suite (section 3.3) ainsi que lors de tests expérimentaux (section 3.5).

1.3 Formulation du problème

Cette section introduit les notations utilisées tout au long de ce chapitre et formalise le problème de l'authentification multimodale d'identité.

Soit N experts distincts et z le vecteur des scores observés lors d'un accès particulier:

$$z = \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ \dots \\ z^{(N)} \end{bmatrix} \tag{1.1}$$

où $z^{(i)}$ représente le score obtenu au droit de l'expert i. Par *score*, nous désignerons toute mesure directe ou indirecte de la probabilité que l'identité du candidat soit celle prétendue être, un score élevé présageant la présence d'un client, un score faible, celle d'un imposteur[4]. Par la suite, nous ferons usage de scores normalisés entre [0,1] où 0 dénote un rejet absolu (imposteur) et 1 une acceptation certaine (client).

En toute généralité, la répartition de ces scores dépend

– du type d'accès que l'on teste, noté $A = \{c, i\}$ où c représente un

3. On suppose un seuil d'acceptation de 0.5.

4. Ce score évolue donc de façon opposée aux distances résiduelles résultantes des mises en correspondances des contours des profils (distance du chanfrein) ou des niveaux de gris (EQM).

accès client et i un accès imposteur;

– de l'identité sous laquelle se présente le candidat, notée Y.

La fonction de distribution des scores relative à une identité (présumée) donnée, conditionnée à un type d'accès particulier, s'écrira dès lors:

$$T_Y(z|A) \tag{1.2}$$

A partir de z, nous devons décider si le candidat est un client ou un imposteur. Soit E_Y le domaine des z amenant à l'acceptation d'un candidat qui se présente sous l'identité Y, et $\overline{E_Y}$ le domaine complémentaire conduisant au rejet du candidat. Nous pouvons alors définir les taux de fausse acceptation (TFA) et de faux rejet (TFR) du superviseur comme suit:

$$\mathsf{TFA}_Y = \int_{E_Y} T_Y(z|A=i)\,dz \tag{1.3}$$

$$\mathsf{TFR}_Y = \int_{\overline{E_Y}} T_Y(z|A=c)\,dz \tag{1.4}$$

Ceux-ci sont illustrés à la figure 1.1.

Par souci de simplification et pour ne pas alourdir les notations qui suivront, l'indice Y sera dorénavant omis. Il ne faudra cependant pas perdre de vue que le domaine d'acceptation E et les valeurs de TFA, TFR dépendent de l'identité sous laquelle se présente le candidat, à moins que l'on n'ait décidé délibérément de modéliser la répartition des scores de façon globale, par une densité $T(z|A)$ caractérisant l'ensemble des clients ou des imposteurs. Dans ce dernier cas, les seuils d'acceptation obtenus par la suite ne seront plus des seuils individuels, propres à chaque client, mais globaux (un seuil unique pour l'ensemble des clients). Notons que l'usage de seuils individuels requiert un nombre de données d'apprentissage élevé pour pouvoir caractériser les distributions $T_Y(z|A)$ propres à chaque client Y. Ne disposant pas, par la suite, de données d'apprentissage en suffisance (cfr. taille relativement petite de la base de données M2VTS), nous ferons généralement usage de seuils globaux.

Comme autre simplification, mineure celle-ci, les notations $A = c$ ou $A = i$ seront par la suite abrégées en c et i, respectivement.

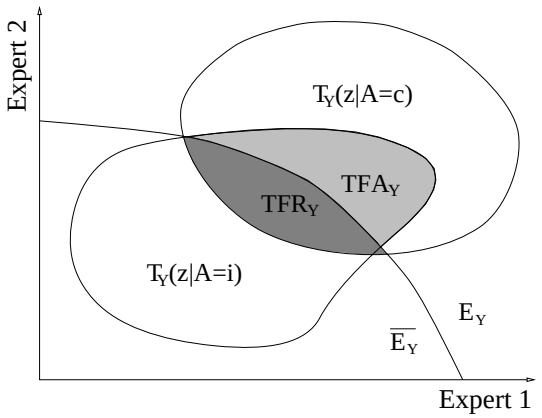


Figure 1.1 - *Répartitions des scores clients et imposteurs conditionnées à un client Y particulier, ainsi que la frontière de décision client/imposteur relative à ce client.*

Après ces changements de notation, les équations (1.3) et (1.4) prennent la forme suivante:

$$\mathsf{TFA} = \int_E T(z|i)\,dz \tag{1.5}$$

$$\mathsf{TFR} = \int_{\overline{E}} T(z|c)\,dz \tag{1.6}$$

Par extension, on peut également introduire les notions de *Taux de Vraie Acceptation* (TVA) et de *Vrai Rejet* (TVR):

$$\mathsf{TVA} = \int_E T(z|c)\,dz \tag{1.7}$$

$$\mathsf{TVR} = \int_{\overline{E}} T(z|i)\,dz \tag{1.8}$$

Cherchons à présent à déterminer le domaine d'acceptation E et par conséquent son complémentaire, le domaine de rejet $\overline{E}$.

1.4 Approche de Neyman-Pearson

L'approche de Neyman-Pearson [62] consiste à minimiser le taux de faux rejet associé à un taux de fausse acceptation α donné, soit:

$$\begin{aligned} \text{minimiser} \quad & \int_{\overline{E}} T(z|c)\, dz \\ \text{sous la contrainte} \quad & \int_{E} T(z|i)\, dz = \alpha. \end{aligned} \tag{1.9}$$

La solution de ce problème est obtenue en utilisant les multiplicateurs de Lagrange, soit en minimisant:

$$L = \int_{\overline{E}} T(z|c)\, dz + \lambda \left[\int_{E} T(z|i)\, dz - \alpha \right] \tag{1.10}$$

où λ est défini positif. Comme

$$\int_{\overline{E}} T(z|c)\, dz = 1 - \int_{E} T(z|c)\, dz, \tag{1.11}$$

nous pouvons réécrire la fonction (1.10) en veillant à travailler à l'intérieur du domaine E exclusivement, soit:

$$L = 1 - \alpha\lambda + \int_{E} \left[\lambda T(z|i) - T(z|c) \right]\, dz. \tag{1.12}$$

Pour que L soit minimale, l'intégrale présente dans le membre de droite doit être la plus négative possible. Ceci revient à inclure dans le domaine d'acceptation E l'ensemble des scores z tels que:

$$\lambda\, T(z|i) - T(z|c) < 0 \tag{1.13}$$

Ceci revient encore à accepter le candidat si le score obtenu vérifie la relation:

$$\frac{T(z|c)}{T(z|i)} > \lambda, \tag{1.14}$$

le facteur λ étant implicitement fixé par la contraite (1.9).

Telle est la règle de décision de Neyman-Pearson, basée sur un rapport des vraisemblances client/imposteur. Il est intéressant de noter que le rapport (1.14) reste optimal pour d'autres critères que celui utilisé par Neyman-Pearson (minimisation du faux rejet à taux de fausse acceptation fixé). Pour illustrer ce propos, quelques critères utilisés en pratique sont envisagés ci-dessous:

– *Minimisation de la fausse acceptation à taux de faux rejet fixé.* Il s'agit du critère dual de celui utilisé par Neyman-Pearson. Dans ce cas, le problème se formule comme suit:

$$\begin{aligned} \textit{minimiser} \quad & \int_E T(z|i)\,dz \\ \textit{sous la contrainte} \quad & \int_{\overline{E}} T(z|c)\,dz = \alpha. \end{aligned} \tag{1.15}$$

La fonction de Lagrange s'écrit

$$L = \int_E T(z|i)\,dz + \lambda \left[\int_{\overline{E}} T(z|c)\,dz - \alpha \right] \tag{1.16}$$

avec λ positif. Le domaine d'acceptation E solution du problème s'exprime comme étant l'ensemble des scores z tels que

$$\frac{T(z|c)}{T(z|i)} > \frac{1}{\lambda} \tag{1.17}$$

– *Minimisation du taux d'égale erreur.* Ceci revient à

$$\begin{aligned} \textit{minimiser} \quad & \int_{\overline{E}} T(z|c)\,dz \\ \textit{sous la contrainte} \quad & \int_{\overline{E}} T(z|c)\,dz = \int_E T(z|i)\,dz, \end{aligned} \tag{1.18}$$

La fonction de Lagrange s'écrit

$$L = \int_{\overline{E}} T(z|c)\,dz + \lambda \left[\int_E T(z|i)\,dz - \int_{\overline{E}} T(z|c)\,dz \right] \tag{1.19}$$

avec λ positif. Le domaine d'acceptation E solution du problème s'exprime comme étant l'ensemble des scores z tels que

$$\frac{T(z|c)}{T(z|i)} > \frac{\lambda}{\lambda - 1} \tag{1.20}$$

– *Minimiser l'erreur totale (maximisation du taux de succès).* Ceci revient à

$$\textit{minimiser} \quad \int_{\overline{E}} T(z|c)\,dz + \int_E T(z|i)\,dz \tag{1.21}$$

sans autre contrainte. L'erreur minimale est obtenue pour le domaine d'acceptation défini par

$$\frac{T(z|c)}{T(z|i)} > 1 \tag{1.22}$$

En conclusion, l'approche de Neyman-Pearson reste optimale pour l'ensemble des critères habituels. Tous débouchent sur un *rapport de vraisemblance* dont seul le niveau d'acceptation varie. L'approche bayesienne développée dans la section suivante nous permettra de mieux cerner la nature de ce seuil.

1.5 Approche de Bayes

Le formalisme de Bayes se base sur deux hypothèses [62]. La première hypothèse fixe a priori les probabilités de se trouver face à un client ou un imposteur (et suppose donc que ces accès sont de nature probabiliste). Soit π_c et π_i ces probabilités respectives ($\pi_c + \pi_i = 1$). La seconde hypothèse associe un coût aux diverses décisions que l'on peut prendre en présence d'un client ou d'un imposteur. Ces coûts sont au nombre de quatre:

– $C_{c|i}$, la pénalité que l'on encourt de décider "client" dans le cas d'un accès imposteur (fausse acceptation)

- $C_{i|c}$, la pénalité que l'on encourt de rejeter un client (faux rejet)
- $C_{c|c}$, le coût lié à l'acceptation d'un client
- $C_{i|i}$, le coût lié au rejet d'un imposteur

L'approche bayesienne revient à minimiser un risque R qui pondère les divers coûts par la probabilité de l'événement auquel ils se rapportent:

$$R = \pi_c \left[C_{c|c}\,\mathsf{TVA} + C_{i|c}\,\mathsf{TFR}\right] \;+\; \pi_i \left[C_{i|i}\,\mathsf{TVR} + C_{c|i}\,\mathsf{TFA}\right] \tag{1.23}$$

Dans le cadre du problème qui nous intéresse, il est logique de ne pas pénaliser l'acceptation d'un client ou le rejet d'un imposteur. Nous fixerons dorénavant $C_{c|c} = C_{i|i} = 0$.

$$R = \pi_c\, C_{i|c}\,\mathsf{TFR} + \pi_i\, C_{c|i}\,\mathsf{TFA} \tag{1.24}$$

En faisant usage des équations (1.5) et (1.6) puis en regroupant les deux intégrales par un jeu d'écriture similaire à celui adopté dans la section 1.4, on montre que le risque minimal est obtenu en acceptant tous les candidats dont le score z satisfait la condition

$$\frac{T(z|c)}{T(z|i)} > \frac{\pi_i\; C_{c|i}}{\pi_c\; C_{i|c}} \tag{1.25}$$

Il s'agit à nouveau d'un rapport de vraisemblance, mais dont le seuil est explicité en termes de probabilités a priori et de coûts d'erreur [18]. Ainsi, plus le risque d'imposture est grand et plus le coût d'une fausse acceptation est élevé comparativement au coût d'un faux rejet, plus le seuil d'acceptation sera élevé.

En pratique néanmoins, il est peu commode de devoir se fixer des probabilités π_i et π_c a priori. Il en est de même des coûts liés aux différents types d'erreur, difficiles à évaluer. C'est pourquoi il est plus aisé de travailler avec une règle de décision générale telle que

$$\frac{T(z|c)}{T(z|i)} > k \tag{1.26}$$

où k est fixé selon les besoins de l'application (on a vu dans la section précédente comment on peut, en faisant varier ce seuil, optimiser de nombreux critères).

L'importance de l'équation (1.25) ne réside pas dans la nature du seuil d'acceptation, mais bien dans l'allure de la règle de décision proposée: un rapport de vraisemblance client/imposteur qui se révèle fournir le critère de décision optimal tant dans le cadre d'une approche de Bayes que de Neyman-Pearson.

Les deux sections qui suivent caractérisent les différents types de superviseurs étudiés par la suite, en faisant usage du formalisme qui vient d'être introduit.

1.6 Fusion dure

Dans le cadre d'une fusion dure (voir section 1.2), le score z observé est un vecteur discret binaire (où 0 désigne un rejet et 1, une acceptation) et les densités de probabilités T deviennent des probabilités P. La règle de décision (1.26) se réécrit sous la forme:

$$\frac{P(z|c)}{P(z|i)} > k \tag{1.27}$$

Explicitons à présent ces probabilités conditionnelles. Le problème de l'authentification d'identité peut se formuler de façon semblable à celui de l'optimisation d'une chaîne de transmission en télécommunications (figure 1.2). Soit deux mots codes possibles définissant la "source":

- le premier mot est associé à la présence d'un imposteur. Il est formé de N zéros et exprime le rejet souhaité par l'ensemble des N experts;
- le second mot correspond à l'accès d'un client et consiste en un vecteur unitaire de taille N.

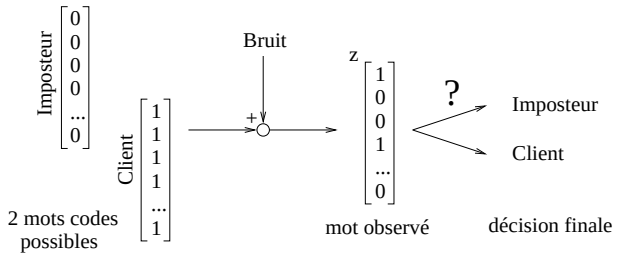


Figure 1.2 - *Modélisation du problème de la fusion dure (analogie avec une chaîne de transmission)*

Après que les experts aient rendu leur avis, on observe un score z qui peut être considéré comme l'un des deux mots codes source entaché d'une certaine erreur (ou d'un "bruit" si l'on reprend l'analogie avec une ligne de transmission). Accepter ou rejeter le candidat revient à déterminer le mot code dont z est issu.

Définissons $p^{(i)}_{d|A}$, la probabilité que l'expert i prenne la décision d en présence d'un accès A avec $d = \{0, 1\}$ où 0 représente un rejet et 1 une acceptation et $A = \{c, i\}$ dénote la présence d'un client ou d'un imposteur. En supposant des experts indépendants[5], la probabilité d'observer z en présence d'un client est donnée par

$$P[z|c] = \prod_{i=1}^{N} p^{(i)}_{d|c} \tag{1.28}$$

En distinguant les N^c_v experts qui ont effectivement accepté le client, des N^c_f experts qui se trompent ($N^c_v + N^c_f = N$), on peut réécrire (1.28) sous la forme

$$P[z|c] \;=\; \prod_{i=1}^{N^c_v} p^{(i)}_{1|c} \prod_{j=1}^{N^c_f} p^{(j)}_{0|c} \;=\; \prod_{i=1}^{N^c_v} \mathsf{TVA}^{(i)} \prod_{j=1}^{N^c_f} \mathsf{TFR}^{(j)} \tag{1.29}$$

où $\mathsf{TVA}^{(i)}$ et $\mathsf{TFR}^{(i)}$ représentent les TVA et TFR associés à l'expert i.

5. Nous reviendrons sur cette hypothèse dans le courant du chapitre 2.

Dans le cas d'un accès imposteur, nous aurons

$$P[z|i] \;=\; \prod_{i=1}^{N_v^i} p_{0|i}^{(i)} \prod_{j=1}^{N_f^i} p_{1|i}^{(j)} \;=\; \prod_{i=1}^{N_v^i} \mathsf{TVR}^{(i)} \prod_{j=1}^{N_f^i} \mathsf{TFA}^{(j)} \tag{1.30}$$

où l'on a également fait une distinction entre les N_v^i experts qui ont effectivement rejeté l'imposteur et les N_f^i experts qui l'ont accepté ($N_v^i + N_f^i = N$).

Notons que les N_v^c, N_f^c, N_v^i et N_f^i dépendent de z et de A. Les $\mathsf{TVA}^{(i)}$, $\mathsf{TFR}^{(j)}$, $\mathsf{TVR}^{(i)}$ et $\mathsf{TFA}^{(j)}$ quant à eux dépendent des seuils d'acceptation librement fixés au droit de chacun des experts[6]. En général, ceux-ci seront dimensionnés durant la phase d'apprentissage du superviseur, comme dans le cas du *superviseur exhaustif dur* étudié au chapitre 4, section 4.4. Une fois les seuil fixés, les $\mathsf{TVA}^{(i)}$, $\mathsf{TFR}^{(j)}$, $\mathsf{TVR}^{(i)}$ et $\mathsf{TFA}^{(j)}$ peuvent alors être évalués sur l'ensemble d'apprentissage.

Les équations (1.29) et (1.30) permettent d'évaluer la règle de décision (1.27). D'un point de vue théorique, cette règle présente une utilité certaine dans le cadre de la fusion de trois experts ou plus. En présence de deux experts seulement, cette règle se rapproche de la fusion logique de type ET ou OU, comme nous le montrons par la suite.

D'un point de vue pratique, le problème peut devenir rapidement complexe à traiter selon le nombre d'experts que l'on désire fusionner. En cause, le libre choix des points opératoires ($\mathsf{TFA}^{(j)}, \mathsf{TFR}^{(j)}$) et donc la difficulté de trouver la combinaison optimale des seuils d'acceptation au droit de chaque expert. Celle-ci est généralement obtenue en effectuant une recherche exhaustive sur ces seuils, lors d'un apprentissage préliminaire. Cette recherche peut rapidement devenir lourde à gérer et ne garantit pas nécessairement de bons résultats (voir phénomène de surapprentissage, section 4.4). C'est pourquoi nous restreindrons la portée de ce travail aux règles simples de fusion dure et plus particulièrement aux opérateurs logiques ET et OU, introduits ci-dessous.

En présence de deux experts de type durs, les mots observés sont au nom-

6. Pour rappel, dans le cadre d'un schéma de fusion dure, les experts transmettent au superviseur une décision et non un score. Cette décision résulte d'un seuillage effectué sur le score qui découle de la mise en correspondance de la personne candidate avec le modèle de référence.

bre de quatre: [0 0], [0 1], [1 0] et [1 1]. Raisonnablement, le candidat sera rejeté si $z = [0\ 0]$ et accepté si $z = [1\ 1]$. Seuls les cas [01] et [10] peuvent poser problème. *En supposant les deux modalités de même fiabilité* ($\forall d, A\ :\ p^{(1)}_{d|A} = p^{(2)}_{d|A}$), ces deux cas seront rattachés à la même décision:

- soit accepter l'individu (minimisation du faux rejet). Ceci revient à fusionner les deux experts par un opérateur logique de type OU;
- soit le rejeter (minimisation de la fausse acceptation). Ceci équivaut à une fusion de type ET.

Ces deux superviseurs logiques seront étudiés aux chapitres 3 et 4, sections 3.2 et 4.4 respectivement.

En présence de deux experts caractérisés par des fiabilités différentes ($\forall d, A\ :\ p^{(1)}_{d|A} \neq p^{(2)}_{d|A}$), il est possible que l'un des cas critiques soit accepté, par exemple [0 1], tandis que l'autre rejeté, soit [1 0]. On remarque dès lors que seul un des deux experts émet un avis pertinent (le deuxième expert dans l'exemple précédent), et que quel que soit l'avis de l'autre expert, il n'est jamais pris en compte. Dans un tel cas, nous ne pouvons plus parler de fusion de données.

Enfin, on notera que la règle logique du *vote majoritaire* souvent utilisée dans le cadre de la fusion de plus de deux experts durs, est équivalente à la règle (1.27) sous l'hypothèse $\forall d, i, j, A\ :\ p^{(i)}_{d|A} = p^{(j)}_{d|A}$.

1.7 Fusion douce

Dans le cadre d'une fusion de type douce, l'usage de la règle (1.26) nécessite l'évaluation des densités de probabilités $T(z|c)$ et $T(z|i)$.

Une première approche revient à évaluer ces densités sur base de densités réelles, acquises durant la phase d'apprentissage du superviseur. En pratique, les scores obtenus durant cet apprentissage servent à construire un histogramme représentatif de chacune des deux classes d'individus (client/imposteur). Cet histogramme fournira par la suite, lors du test, une estimation de $P(z|A)$. La règle (1.26) s'interprète alors comme un rapport de probabilités $P(z|c)$ et $P(z|i)$ et non de densités. Pour être fiable, cette méthodologie nécessite le recours à un nombre élevé de données d'ap-

prentissage, ce qui s'avèrera impossible dans la suite de ce travail. C'est pourquoi la méthode suggérée ici n'a pu être mise en œuvre.

La seconde approche consiste à faire l'hypothèse de distributions $T(z|A)$ particulières, généralement exprimées de façon analytique (approche paramétrique). En remplaçant les densités $T(z|c)$ et $T(z|i)$ par leur expression, on peut alors dériver une règle de décision qui sera optimale lorsque les densités observées satisferont l'hypothèse faite. Le superviseur statistique de Fisher introduit au chapitre 4, section 4.5 illustre cette approche et se base sur une approximation gaussienne. Cette approximation débouche sur une règle de décision linéaire ou quadratique selon que l'on travaille avec une matrice de covariance entre modalités identique aux clients et imposteurs, ou spécifique à chaque classe.

Enfin, rien ne nous impose de suivre la règle du rapport de vraisemblance, et dans le cas d'une règle de décision linéaire comme celle obtenue ci-dessus en se basant sur une distribution gaussienne, on peut se demander si il n'existe pas de meilleure droite, correspondant davantage aux densités de probabilité réelles. Une telle démarche sera proposée dans le cadre des chapitre 3, section 3.3 et chapitre 4, section 4.4.

Chapitre 2

Approche graphique

2.1 Contexte

Après avoir introduit la fusion de données, et en particulier l'authentification multimodale d'identité, de façon mathématique, ce chapitre privilégiera une approche illustrée et graphique du problème (sections 2.2 et 2.3). Il tentera également de caractériser les propriétés dont doivent jouir les experts afin d'obtenir un gain en performance optimal lors de la fusion de ceux-ci (section 2.4). Ces considérations émises, nous analyserons les caractéristiques relatives aux experts développés dans le cadre de ce travail (section 2.5). Ensuite, nous analyserons l'hypothèse d'indépendance entre experts utilisés, hypothèse déjà utilisée dans le chapitre précédent, et qui nous permettra en outre de simplifier de façon considérable certains développements théoriques ultérieurs (section 2.6).

2.2 Représentation du problème

Tout au long des chapitres précédents, nous avons eu l'occasion d'introduire les concepts de fausse acceptation et de faux rejet. En ne se limitant qu'à un seul expert, ces taux peuvent être représentés de façon graphique à la figure 2.1, où $T(x|c)$ et $T(x|i)$ désignent respectivement les répartitions des scores clients et imposteurs, conformément aux notations utilisées dans le

cadre du chapitre précédent, et k un seuil d'acceptation donné[1].

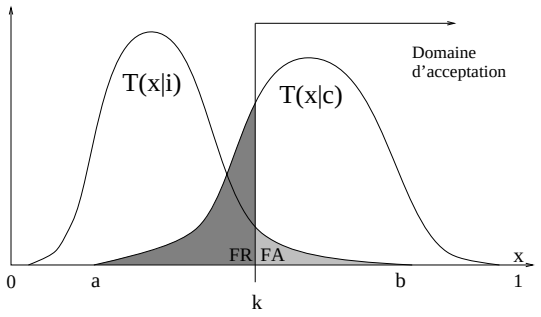


Figure 2.1 - *Répartition des scores clients $T(x|c)$ et imposteurs $T(x|i)$, FA et FR*

On remarquera sur la figure 2.1 que rares sont les cas où nous pouvons déterminer avec certitude la présence d'un client ou d'un imposteur (scores supérieurs à b ou inférieurs à a). Il existe en effet une large plage de valeurs où le score obtenu pourrait être tant celui d'un client que celui d'un imposteur. Aussi, voit-on clairement apparaître le problème intrinsèque en théorie de la décision, à savoir l'impossibilité de minimiser les TFA et TFR simultanément (ce qui reviendrait à respectivement augmenter et diminuer le seuil d'acceptation). Nous veillerons donc à travailler avec un critère global, tel celui d'une erreur totale pondérée par exemple:

$$\alpha\,\mathsf{TFA} + (1-\alpha)\,\mathsf{TFR} \tag{2.1}$$

où α est un paramètre variant de 0 à 1 selon l'application envisagée: il sera d'autant plus grand qu'il est important de rejeter un maximum d'imposteurs, au risque de devoir rejeter plus de clients également.

Représenter deux courbes distinctes $T(x|c)$ et $T(x|i)$ telles celles de la figure 2.1, suppose implicitement que l'on dispose d'une information a priori sur la classe à laquelle appartient l'individu dont on vérifie l'identité (client/imposteur). Bien entendu, cette information n'est pas disponible dans un système opérationnel, puisqu'il s'agit de l'inconnue que l'on cherche

1. Pour rappel, k désigne à présent un seuil sur les scores et non sur les distances résiduelles comme dans la première partie de ce travail.

précisément à déterminer. C'est seulement durant une phase dite d'apprentissage, phase de simulation durant laquelle chaque expert est entraîné à pouvoir dissocier un client d'un imposteur, que l'on connaît effectivement la nature du test que l'on réalise (voir protocole de test, chapitre 3, partie 1). C'est pendant cette phase donc que nous pourrons explicitement calculer les TFA et TFR et optimiser ainsi le seuil d'acceptation k selon le critère de performance choisi. On espère alors que l'ensemble d'apprentissage utilisé soit représentatif des cas qui seront rencontrés une fois le système d'authentification opérationnel et que le seuil d'acceptation fixé fournisse, dans la pratique, les performances attendues. L'étude de la sensibilité des performances d'un système d'authentification aux ensembles d'entraînement et de test fera l'objet du chapitre 4.

L'équivalent bidimensionnel de la figure 2.1 est représenté à la figure 2.2. A chaque expert est associé un axe différent sur lequel sont reportés les distributions de scores clients et imposteurs relatives à cet expert. Ces distributions individuelles servent alors à tracer des courbes d'iso-densités sur l'ensemble du plan (traits interrompus). Afin de simplifier cette représentation, seule une iso-densité par classe d'individu sera représentée par la suite (en trait continu). Celle-ci sera choisie suffisamment large, c'est à dire associée à un percentile suffisamment élevé[2] afin de pouvoir délimiter au mieux les espaces occupés en majorité par des scores imposteurs et ceux occupés par des scores clients.

Si dans un schéma monomodal, la décision d'accepter ou de rejeter un individu revenait à effectuer un simple seuillage (global ou individuel) sur le score obtenu au sein de la modalité unique, un schéma multimodal ouvre quant à lui une multitude de possibilités nouvelles pour pouvoir distinguer un client d'un imposteur. En adoptant la représentation de la figure 2.2, le problème de l'authentification de personnes peut se ramener à celui de partitionner l'espace bidimensionnel des scores en deux sous-ensembles disjoints, celui des clients et des imposteurs. Un individu sera alors accepté comme client si son score appartient au sous-ensemble associé, et rejeté dans le cas contraire (cfr. les sous-ensembles E et $\overline{E}$ définis au chapitre 1). Contrairement à l'unique technique de seuillage dont nous disposions en monomodal, il existe de nombreuses possibilités de définir ces sous-ensembles et de caractériser la frontière qui les sépare. Quelques exemples de classifications bidimensionnelles sont repris à la figure 2.3. La première sous-figure repré-

2. Un percentile associé à une courbe de niveau, dénombre le pourcentage d'individus (de la même classe) localisés à l'intérieur de la courbe.

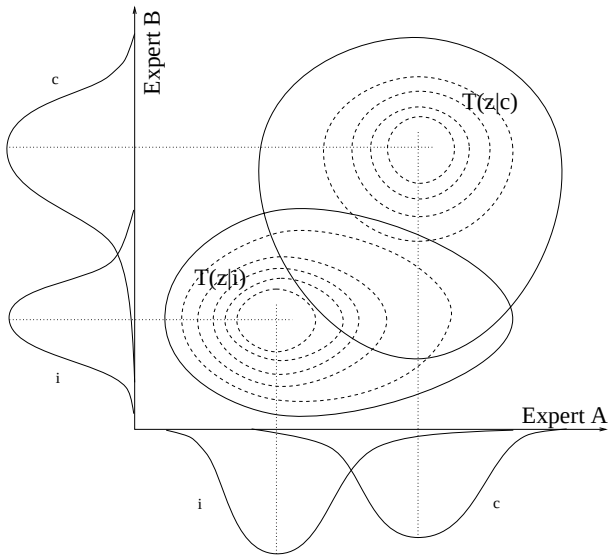
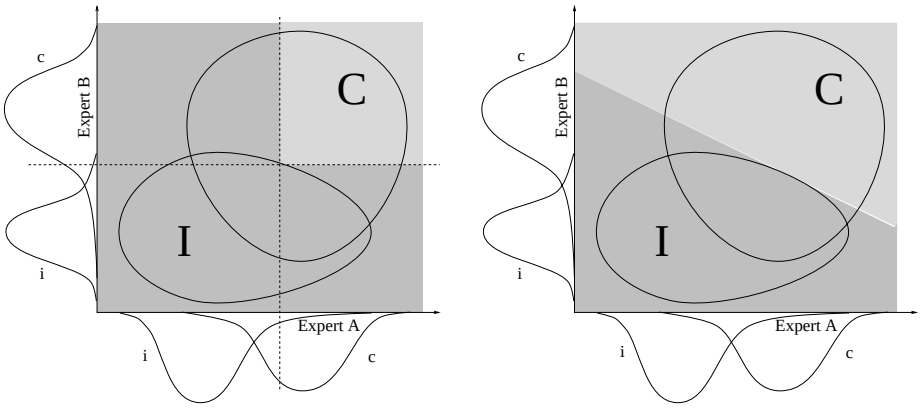


Figure 2.2 - *Représentation bidimensionnelle de la répartition des scores clients et imposteurs dans le cadre de la fusion de deux experts*

sente la classification obtenue en seuillant chaque modalité séparément et en combinant les résultats par un opérateur logique de type ET (un client est accepté si le score obtenu dans *chacune* des modalités est supérieur au seuil considéré). Le deuxième cas représente une frontière de décision associée au seuillage d'une combinaison linéaire des scores issus de chaque modalité. Le troisième cas combine plusieurs seuillages linéaires. Le quatrième et dernier cas est quelque peu fantaisiste mais nous montre la multitude de possibilités qui nous sont offertes!

Tout au long de cette seconde partie, nous tenterons de comparer différentes techniques de fusion entre elles et de caractériser le gain que l'on peut en attendre par rapport aux performances monomodales obtenues dans la première partie de ce travail.



(a) Seuillage individuel + ET

(b) Classification linéaire

(c) Classification polygonale

(d) Classification quelconque

Figure 2.3 - *Classification des scores clients/imposteurs issus de deux experts (fusion bimodale)*

2.3 Notion d'indépendance

La figure 2.2 représente en fait deux experts indépendants l'un de l'autre, c'est-à-dire des experts pour lesquels le score obtenu par l'un ne dépend en rien du score obtenu par l'autre. Ceci revient encore à dire que la probabilité d'observer un score combiné $z = [z^{(1)}\ z^{(2)}\ ...\ z^{(n)}]$ est obtenue en effectuant le produit des probabilités individuelles pour chaque l'expert, soit, dans le cas continu:

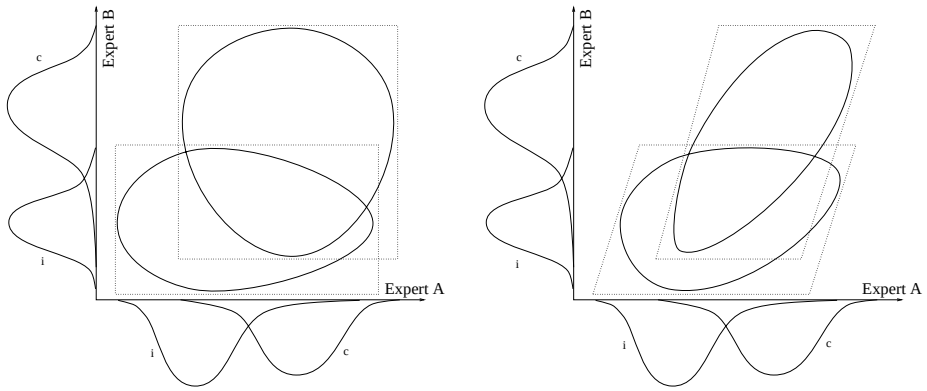
$$T[z|A] \ = \ \prod_{i=1}^{N} T^{(i)}[z^{(i)}|A] \tag{2.2}$$

L'application de l'équation (2.2) permet de calculer les courbes d'iso-densité représentées à la figure 2.2. Si les experts n'avaient pas été indépendants, des probabilités plus élevées que celles données par l'équation (2.2) auraient été observées aux endroits où le score d'un expert laissait présager de celui de l'autre. Un exemple de ce cas particulier est illustré à la figure 2.4. La sous-figure (a) reprend le cas d'experts indépendants et représente les iso-densités caractéristiques des clients et imposteurs (traits continus) ainsi que les domaines au sein desquels se répartissent la *totalité* des scores clients et imposteurs (pointillés)[3]. Dans le cas d'experts indépendants, ces derniers ont nécessairement une forme rectangulaire. La sous-figure (b) introduit une dépendance positive entre les deux experts, à savoir qu'un score élevé dans une modalité laisse présager un score élevé dans la seconde modalité également. Ceci se remarque particulièrement par la forme parallèlipipédique des domaines des scores repris en pointillés (un score élevé obtenu par un expert supprime la possibilité d'obtenir un score faible dans l'autre). Notons que la présence de domaines non rectangulaires traduit nécessairement une dépendance entre experts, mais que la réciproque ne se vérifie pas (une dépendance particulière entre experts peut aussi engendrer des domaines rectangulaires).

Alors que l'équation 2.2 définit clairement la notion d'indépendance entre experts, cette notion peut néanmoins prêter à confusion. Intuitivement, nous interprétons la propriété d'indépendance comme étant *l'absence de toute interaction.* Ainsi, un expert vocal et un expert image pourraient, par exemple, être considérés comme indépendants[4]. Cette interprétation de ce

3. Ces domaines correspondent aux percentiles 100%.

4. pour autant que l'on suppose que la forme du visage n'a aucune influence sur la



(a) Experts indépendants (b) Experts dépendants

Figure 2.4 - *Distribution des scores clients/imposteurs : influence de la dépendance*

qu'est l'indépendance comme définie en 2.2 est correcte. Une certaine confusion peut néanmoins apparaître lorsque l'on remarque que quelle que soit la modalité considérée, l'authentification d'un client engendrera en moyenne de meilleurs scores que celle d'un imposteur. Ainsi, il suffit qu'un expert authentifie un client et offre un score élevé pour laisser présager d'un score élevé pour les autres experts également. La corrélation que l'on devine ici, n'est en rien liée à la dépendance qui peut exister entre nos différents experts[5]. Elle traduit en fait la convention implicite qui consiste, pour chaque expert, à associer un score élevé à une bonne mise en correspondance.

Pour se départir de cette ambiguïté possible, mieux vaut ne pas raisonner avec les scores issus des différentes modalités, mais avec les erreurs qu'elles commettent, soit $\epsilon = |z|$ pour un client et $\epsilon = |1 - z|$ pour un imposteur.

voix d'un individu

5. Pour s'en convaincre, il suffit de se placer dans un cas extrême et de supposer différents experts *indépendants* et parfaits, ne souffrant d'aucune fausse acceptation ni de faux rejet. Nous noterions alors à tout moment des scores et des décisions rigoureusement identiques pour chacun d'eux, à savoir l'acceptation dans le cas d'un accès client (score égal à l'unité) ou le rejet d'un imposteur (score nul). La forte corrélation que l'on note ici mesure simplement la dépendance qui existe entre la classe d'un individu (client/imposteur) et le résultat de l'authentification qui en résulte (accès/rejet).

Deux modalités seront alors corrélées, si les erreurs qu'elles commettent sont corrélées.

2.4 Experts optimaux

Une idée largement répandue dans le domaine de la fusion de données, est de lier le gain en performance issu de la fusion, à l'indépendance des données que l'on fusionne [65]. Ainsi, pourrait-on croire que pour pouvoir bénéficier de performances accrues lors de la fusion de deux experts, il est nécessaire que ceux-ci *soient les moins corrélés possible*. Si dépendance il y avait, l'information relative à l'identité d'une personne fournie par l'une des modalités serait en partie redondante avec celle fournie par l'autre, ce qui se traduirait par un manque à gagner lors de la fusion. C'est du moins ce que l'on pourrait croire a priori.

La figure 2.5 illustre différents cas de fusion effectués sur des experts caractérisés par des répartitions de scores uniformes afin d'en faciliter la représentation. Sur cette figure sont représentés, dans trois cas différents, les domaines des scores clients et imposteurs ainsi que les seuils d'acceptation et les frontières de décision associés aux points de fonctionnement caractérisés par la condition TFA=TFR. Dans le premier cas (a), on voit comment la fusion améliore les performances du système entier: de deux modalités indépendantes souffrant d'un TEE de 25%, on réduit ce taux de moitié par l'utilisation d'une frontière de décision linéaire[6]. Le cas (b) introduit une *corrélation positive*[7] entre les différents experts. On remarque comment cette corrélation positive porte préjudice à la bonne séparation des nuages de points client et imposteur: des points qui se trouvaient en dehors de l'intersection des deux classes dans le cas (a) peuvent à présent se retrouver à l'intérieur de celle-ci. Ceci entraîne inévitablement une dégradation des performances après fusion par rapport au cas d'experts indépendants. Cela ne veut pas dire pour autant que toute dépendance est néfaste. Ainsi, dans le cas (c), une *dépendance négative*[8] engendre l'effet contraire et aide, comme on le verra par la suite, à pouvoir mieux séparer les clients des imposteurs. Il en résulte dès lors des performances de fusion accrues.

6. Seuillage d'une combinaison linéaire des scores issus des deux experts.

7. Les deux experts éprouvent alors simultanément des difficultés à authentifier un individu.

8. L'accroissement de la difficulté à authentifier un individu pour une modalité correspond à une augmentation de la facilité pour l'autre.

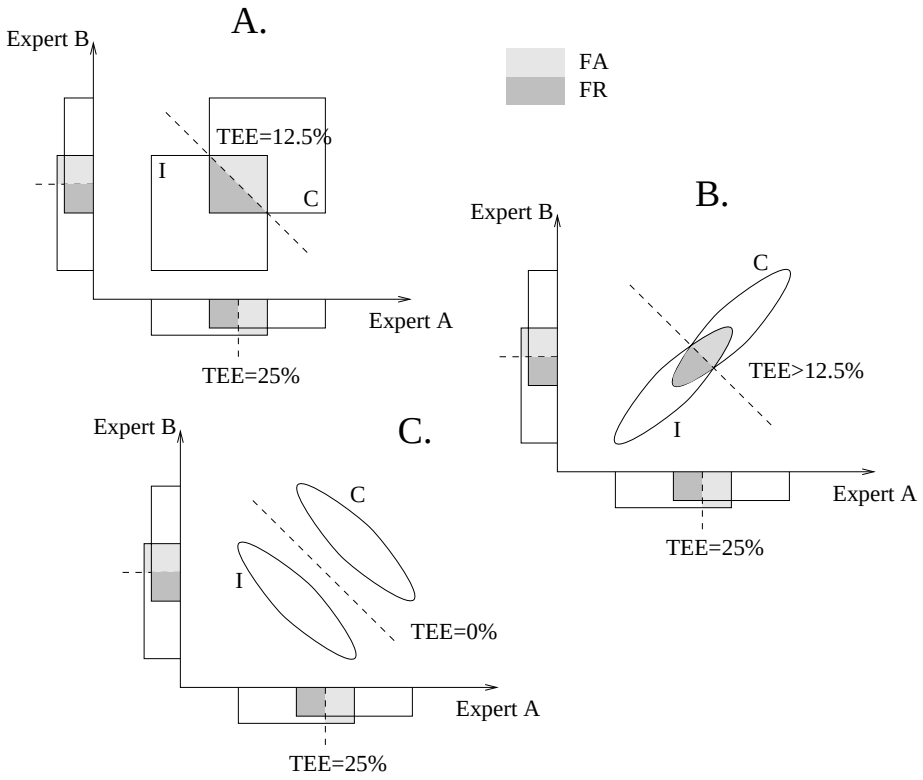


Figure 2.5 - *Influence de la dépendance des experts sur le gain de fusion*

Ainsi les conditions requises pour bénéficier du meilleur gain de fusion ne sont pas forcément liées à l'indépendance des experts, mais plutôt à la séparation qui existe entre les nuages de scores clients et imposteurs. Plus ces nuages seront distincts, meilleures pourront être les performances en aval du superviseur.

Essayons à présent de voir pourquoi une corrélation négative induira toujours une meilleure séparation entre les scores clients et imposteurs.

Soit $\mu_{z|i}^{(i)}$ et $\mu_{z|c}^{(i)}$ les moyennes des scores imposteurs et clients pour une modalité i donnée. Nous avons:

$$\mu_{z|i}^{(i)} < \mu_{z|c}^{(i)} \tag{2.3}$$

Si tel n'était pas le cas, la modalité considérée n'aurait aucun sens puisqu'en principe il est plus facile de mettre en correspondance un client avec son propre modèle qu'un imposteur avec un modèle client. Dans un graphe bidimensionnel tel celui représenté à la figure 2.6, cela se traduit par une population d'imposteurs caractérisée par un centre de gravité localisé plus bas et plus à gauche que celui des clients. Ceci signifie encore que la droite joignant les centres de gravité des nuages de points clients et impoteurs est caractérisée par une pente positive, jamais négative. Elle est illustrée à la figure 2.6.

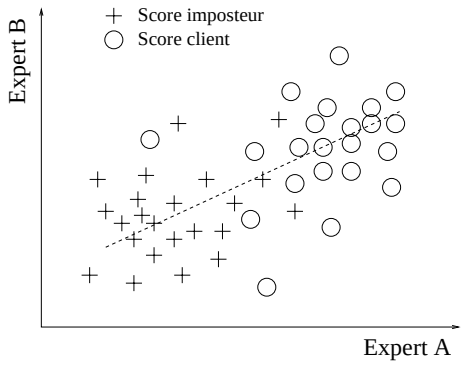
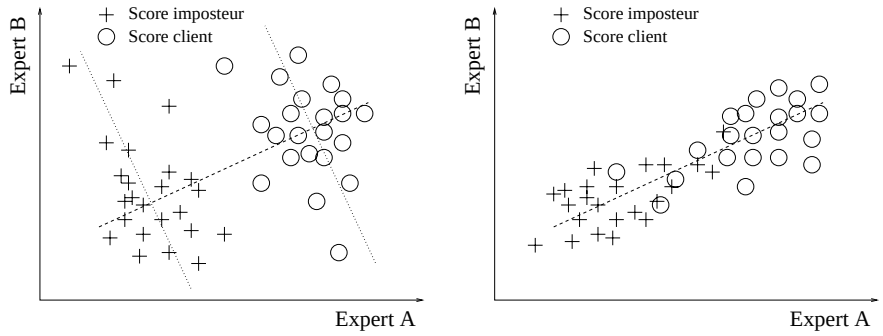


Figure 2.6 - *Répartition bidimensionelle des scores clients et imposteurs*

Si l'on veut introduire une meilleure séparation entre ces nuages, on veillera à disperser les scores appartenant à une même classe d'accès, perpendiculairement à la droite représentée à la figure 2.6. Ceci revient alors à introduire une corrélation négative entre experts (figure 2.7(a)). A l'inverse, une corrélation positive, qui étale les scores dans la même direction que la droite qui joint les moyennes $\mu_{z|i}^{(i)}$ et $\mu_{z|c}^{(i)}$ est préjudiciable à la bonne séparation des clients et des imposteurs (figure 2.7(b)).

Un raisonnement plus intuitif permet d'aboutir aux mêmes conclusions. Ainsi, dans le cas de l'authentification d'un client par l'intermédiaire de la fusion de deux experts, il est souhaitable d'observer les deux scores les plus élevés possibles (cas idéal). Si jamais un des experts ne parvenait pas à authentifier le client correctement, il faut espérer que le second ne soit pas à son tour défaillant et de surcroît qu'il fournisse un score suffisamment bon pour influencer le superviseur à accepter le client. En d'autres mots, il est bénéfique de rechercher, dans une telle situation, une certaine corrélation



(a) Corrélation intra-classe négative (b) Corrélation intra-classe positive

Figure 2.7 - *Répartition bidimensionelle des scores clients et imposteurs: effet d'une corrélation entre modalités*

négative entre experts, puisque l'erreur élevée de l'un doit idéalement être compensée par une erreur suffisamment faible de l'autre. Un raisonnement équivalent peut être également tenu pour les imposteurs. Idéalement, ceux-ci devraient être caractérisés par un score aussi bas que possible dans les deux modalités. Si un imposteur parvient à se faire passer pour un client dans l'une d'elles, nous avons intérêt à ce qu'il soit rejeté sans équivoque dans l'autre. Ceci confirme à nouveau l'utilité d'une forme de corrélation négative entre scores issus de différents experts.

Ainsi voyons-nous apparaître l'intérêt qu'il y a à développer des experts dont les scores sont caractérisés par une dépendance négative. Une telle dépendance peut apparaître dans le cas d'experts utilisant des caractéristiques biométriques mutuellement exclusives. Nous pourrions citer l'exemple d'un expert travaillant sur les yeux ouverts (reconnaissance de l'iris) et un autre sur les yeux fermés (exemple fictif d'une reconnaissance des paupières). Dans le cas où l'on ne sait déterminer a priori si les yeux sont ouverts ou fermés, ces deux modules se complètent parfaitement: les conditions opératoires mettant l'un en défaut (paupières ouvertes ou fermées), garantissent le bon fonctionnement de l'autre. La corrélation apparaissant ici est bénéfique pour le système. Il en est de même, dans une moindre mesure, dans le cas (réel celui-ci) d'experts traitant du visage de face et du visage de profil, lorsqu'ils travaillent sur une même image d'un visage pour lequel l'angle

de vue est indéterminé (cas d'une personne photographiée à son insu, par exemple).

Une autre façon de faire apparaître une dépendance négative, sans toutefois devoir faire usage de caractéristiques biométriques exclusives, consiste à combiner des méthodes d'authentification dont les faiblesses de l'une correspondent précisément aux conditions dans lesquelles l'autre est à même de donner un bon résultat. Ces méthodes d'authentification peuvent à la limite traiter des caractéristiques biométriques identiques. Comme exemple, nous pourrions citer le cas de deux experts travaillant sur le visage de face mais qui travailleraient chacun avec des images relatives à des angles de prise de vues différents (cas de deux caméras désaxées). Ceci permet de bénéficier d'une plus grande souplesse lors du positionnement du client par rapport au système et il suffit que le visage de face soit visible par une des deux caméras seulement pour que le client puisse être authentifié convenablement. Notons néanmoins que l'utilisation d'experts "complémentaires" basés sur des caractéristiques biométriques identiques, en comparaison avec l'usage de caractéristiques exclusives, introduit le risque que le candidat ne puisse être effectivement reconnu par les caractéristiques biométriques traitées (tout à coup le client se présente avec une barbe, par exemple). Dès lors, quoique corrélés négativement, aucun des deux experts ne sera à même de traiter le client, qui verra ainsi son accès refusé.

Terminons par une mise en garde relative aux bienfaits d'une corrélation négative. Si il est vrai qu'une telle corrélation peut introduire une meilleure séparation des nuages client et imposteur, elle peut en contrepartie porter préjudice à la robustesse et à la stabilité du superviseur. Reprenons le cas des figures 2.5 B et C. Pour peu que la position relative des nuages de scores par rapport à la frontière de décision soit modifiée (ce qui peut se produire en passant d'un environnement d'apprentissage aux conditions réelles de test), le TFA ou TFR peut augmenter sensiblement dans le cas d'experts corrélés négativement. Le problème sera d'autant plus critique que les nuages client et imposteur représentés à la figure 2.5 C seront proches l'un de l'autre. Dans le cas d'une corrélation positive, la sensibilité des performances du superviseur par rapport au positionnement de la frontière de décision sera moindre.

2.5 Quid de nos experts?

Attardons-nous à présent à décrire la corrélation qui existe en pratique entre les différents experts introduits durant la première partie de ce travail. Rien ne permet de présager une corrélation négative, potentiellement bénéfique pour le système comme nous l'avons vu précédemment. Nous pouvons seulement espérer qu'il n'y ait pas de corrélation positive trop élevée pour que le superviseur puisse être capable d'améliorer les performances des experts individuels de façon sensible.

Au sens statistique du terme, la corrélation ρ_{ij} se définit comme suit:

$$\rho_{ij} = \frac{E\{(\epsilon_i - \mu_{\epsilon i})(\epsilon_j - \mu_{\epsilon j})\}}{\sigma_{\epsilon i}\ \sigma_{\epsilon j}} \tag{2.4}$$

où $\mu_{\epsilon i}$ et $\sigma_{\epsilon i}$ désignent respectivement la moyenne et l'écart-type des erreurs d'authentification pour une modalité i donnée. On peut montrer que l'indépendance entre modalités i et j implique une corrélation $\rho_{ij} = 0$. La réciproque, quant à elle, n'est pas forcément vraie. Les valeurs de ρ_{ij} pour l'ensemble des modalités et des experts traités dans ce travail sont repris au tableau 2.1. Ces valeurs permettent d'isoler les modalités les plus corrélées, à savoir (par ordre décroissant de corrélation):

- les modalités *Relief du Profil* (MRP) et *Profil Niveaux de Gris* (MPNG). Cette constatation semble assez logique puisque ces modalités sont relatives à la même vue de profil et que les paramètres de compensation issus de la MRP sont directement réutilisés par la MPNG. Dans le cas où ces paramètres ont été mal estimés, une même erreur affectera inévitablement la MRP et la MPNG.

- la MPNG avec l'*Expert Frontal* (EF). Cette corrélation s'explique par le fait que les deux algorithmes se basent sur le calcul d'une distance directement liée à la luminance de l'image et peuvent ainsi être simultanément affectés par de mauvaises conditions d'éclairage.

- l'EF avec l'*Expert Vocal* (EV). L'existence d'une telle corrélation peut paraître étonnante, puisque rien ne permet à première vue de lier la voix et le visage de face. Néanmoins, elle peut s'expliquer par le fait que le contenu fréquentiel du signal de parole varie selon la position

relative du microphone et du locuteur[9], que cette variation est préjudiciable à la bonne reconnaissance du locuteur et que l'EF peut lui aussi éprouver certaines difficultés à devoir authentifier un client lorsque celui-ci ne se positionne pas correctement face à la caméra[10]. Ce préjudice qui se fait ressentir simultanément au sein des experts frontal et vocal, aura pour effet d'introduire une corrélation positive entre ces deux experts. Il ne s'agit donc pas d'une corrélation réelle entre le signal de parole et les caractéristiques du visage de face, mais plutôt d'une inaptitude commune à ces deux experts à traiter certaines conditions opératoires critiques. Cela étant remarqué, nous aurions peut-être pu obtenir une forme de corrélation négative (bénéfique) en évitant l'alignement rigoureux entre la caméra, le microphone et l'individu candidat. On aurait ainsi évité qu'un client puisse se présenter dans une position qui soit défavorable aux experts parole et frontal simultanément.

Les autres combinaisons d'experts ou de modalités offrent une estimation de corrélation suffisamment faible pour pouvoir raisonnablement présager de leur indépendance.

ρ_{ij}	Relief Profil (MRP)	Profil Niveaux Gris (MPNG)	Expert Profil (EP)	Expert Frontal (EF)
Relief Profil (MRP)	-	39.7%	-	-7.6%
Profil Niveaux Gris (MPNG)	39.7%	-	-	22.8%
Expert Profil (EP)	-	-	-	7.6%
Expert Frontal (EF)	-7.6%	22.8%	7.6%	-
Expert Vocal (EV)	-5.4%	5.6%	-2.3%	22.0%

Tableau 2.1 - *Etude de la dépendance entre modalités et experts: valeurs de* ρ_{ij}

9. On remarquera une légère atténuation des composantes basses-fréquences lorsque le locuteur est de biais (ce phénomène peut dépendre du type de microphone utilisé).

10. Les distorsions du visage qui apparaissent alors dans l'image ne peuvent plus être parfaitement compensées par les transformations $\{t_x, t_y, z_x, z_y, \theta_{xy}\}$ introduites à la section 2.2 (voir partie 1), et détérioreront la mise en correspondance du visage candidat avec sa référence.

2.6 Hypothèse d'indépendance des experts utilisés

L'hypothèse de modalités indépendantes, même si elle ne correspond pas strictement à la réalité de nos experts est souvent utilisée dans la littérature car elle permet d'aborder plus simplement de façon théorique le problème de la fusion de données [66, 19, 9]. C'est pour cette raison qu'elle fut introduite dans la section 1.6 et qu'elle sera encore utilisée dans le courant du chapitre 3. Il nous reste à montrer que, pour les experts frontal et vocal et dans une moindre mesure les experts profil et frontal, cette hypothèse semble ne pas porter préjudice à la validité des résultats que nous obtiendrions de la sorte.

Cette section se propose de déterminer un taux de corrélation maximal admissible en dessous duquel l'hypothèse d'indépendance peut être faite. Le sujet étant relativement vaste, nous nous sommes limités à l'étude du cas suivant.

Soit deux experts indépendants A et B fournissant les scores $z^{(a)}$ et $z^{(b)}$ respectivement. Faisons l'hypothèse de distributions de scores gaussiennes, caractérisées par un écart-type commun aux clients et aux imposteurs:

$$\sigma_c^{(a)} = \sigma_i^{(a)} = \sigma_c^{(b)} = \sigma_i^{(b)} \tag{2.5}$$

Dans un tel cas, la règle de décision optimale (1.26) peut se réécrire sous forme linéaire[11] et la frontière de décision qui s'y rapporte est une droite perpendiculaire à celle qui joint les centres des deux gaussiennes client et imposteur. En supposant les centres de ces deux gaussiennes alignés le long de la première bissectrice[12]

$$\mu_c^{(a)} - \mu_i^{(a)} = \mu_c^{(b)} - \mu_i^{(b)} \tag{2.6}$$

la combinaison optimale des scores est alors donnée par:

$$s_{opt} = 0.5\ z^{(a)} + 0.5\ z^{(b)} \tag{2.7}$$

11. Démonstration faite dans le cadre de la section 4.5.

12. Cette hypothèse n'est en rien restrictive et correspond à un choix particulier du système d'axes de coordonnées.

Envisageons maintenant le cas de deux experts dépendants. Pour cela, introduisons une dépendance explicite entre A et B et remplaçons l'expert B de la façon suivante (cas d'une dépendance linéaire):

$$B' = \lambda\ A + (1 - \lambda)\ B \tag{2.8}$$

où λ est un facteur compris entre 0 et 1, qui varie selon la dépendance que l'on veut introduire entre les experts A et B'. En combinant les équations (2.7) et (2.8), la fusion optimale des scores a et b' au sens de (1.26) est alors donnée par:

$$s_{opt} = \frac{1 - 2\lambda}{2 - 2\lambda}\ z^{(a)} + \frac{1}{2 - 2\lambda}\ z^{(b')} \tag{2.9}$$

Si les experts A et B' avaient été supposés indépendants, nous aurions a priori adopté la règle de fusion donnée par l'équation (2.7), alors que seule l'équation (2.9) fournit l'optimum dans ce cas. Ces deux équations permettent donc de chiffrer l'erreur commise lors de l'évaluation des performances optimales d'un système fusionnant deux experts erronément supposés indépendants.

Le tableau 2.2 reprend, pour différents taux de corrélation, les TEE obtenus en fusionnant les deux experts A et B' selon les équations (2.7) et (2.9) dans les colonnes TEE_{indep} et TEE_{dep} respectivement. Outre le cas gaussien envisagé ci-dessus, ce tableau fournit également les résultats relatifs à des distributions de scores uniformes. Compte tenu des hypothèses (2.5) et (2.6), chaque expert jouit des mêmes performances et l'on a fixé arbitrairement le TEE de chacun d'eux à 20%. Les taux de corrélation varient de 0% – soit une indépendance totale, les équations 2.7 et 2.9 fournissant alors une même mesure de performance optimale après fusion – jusqu'à 100%, l'expert B' devenant identique à l'expert A et le TEE obtenu étant alors égal au taux marginal (20%).

L'erreur ξ commise en fusionnant deux experts erronément considérés comme indépendants, a été définie comme suit:

$$\xi = \frac{TEE_{indep} - TEE_{dep}}{TEE_{marginal} - TEE_{dep}} \tag{2.10}$$

Cette erreur est illustrée à la figure 2.8 en fonction du taux de corrélation

$\rho_{ab'}$ (%)	Distributions Uniformes		Distributions Gaussiennes	
	TEE_{indep} (%)	TEE_{dep} (%)	TEE_{indep} (%)	TEE_{dep} (%)
0	8.0	8.0	11.8	11.8
10	8.2	8.0	11.8	11.8
20	8.4	8.0	12.0	11.8
30	8.6	8.0	12.3	11.8
40	8.9	8.0	12.7	11.8
50	9.5	8.0	13.2	11.8
60	10.1	8.0	13.8	11.8
70	11.0	8.0	14.5	11.8
80	12.1	8.0	15.2	11.8
90	14.4	8.0	16.4	11.8
99	18.1	8.0	18.6	11.8
100	20.0	-	20.0	-

Tableau 2.2 - *Etude de l'erreur commise sur l'évaluation des performances optimales d'un système, selon le niveau de corrélation $\rho_{ab'}$ entre experts (chaque expert est caractérisé par un TEE marginal de 20%).*

$\rho_{ab'}$. On y remarque une croissance caractérisée par une allure exponentielle et une courbe relativement stable pour des corrélations inférieures à 50% environ. Ce n'est qu'au-delà de cette limite que l'erreur se met à croître rapidement.

Nous pouvons à présent reprendre les résultats du tableau 2.1, à savoir les niveaux de corrélation mesurés entre nos différents experts et évaluer combien l'hypothèse d'indépendance pourrait être préjudiciable à la suite de notre travail. La corrélation maximale observée entre nos différents experts est obtenue pour la combinaison des experts vocal et frontal, soit un taux de 22%. A ce taux associé au pire des cas correspond une erreur ξ de moins de 3% (tant sous l'hypothèse uniforme que gaussienne). *Pour autant que l'on puisse considérer les experts étudiés de fiabilité égale et que les dépendances qui les lient peuvent être approchées par une relation linéaire, l'hypothèse d'indépendance entre modalités semble donc acceptable et pourra être utilisée dans la suite de ce travail sans risquer de compromettre dangereusement la validité de nos résultats.*

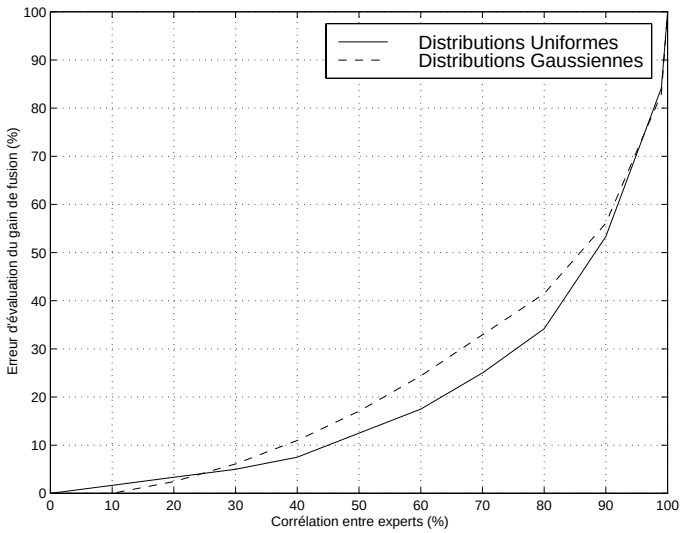


Figure 2.8 - *L'erreur commise en fusionnant deux experts erronément considérés comme indépendants, en fonction du taux de corrélation existant entre les deux experts étudiés.*

Chapitre 3

Analyse a posteriori (scores présupposés connus)

3.1 Contexte

Les deux grandes familles de superviseurs présentées à la section 1.2 seront étudiées et comparées dans les chapitres qui suivent. Comme il est illusoire de vouloir aborder le sujet de façon exhaustive, notre étude fut restreinte aux superviseurs suivants:

- la fusion de décisions fera usage des deux opérateurs logiques les plus intuitifs et les plus répandus, à savoir les opérateurs ET et OU;
- la fusion de scores quant à elle, sera basée sur la meilleure combinaison linéaire des scores issus des différentes modalités.

Dans ce chapitre, l'analyse des performances sera établie *a posteriori*. Une telle analyse revient à comparer les meilleures performances qu'il est possible d'obtenir sur un ensemble de test *donné*, pour chacun des superviseurs étudiés. En d'autres termes, les scores clients et imposteurs sont a priori supposés connus. Ceci permet de calculer les bornes supérieures aux performances qu'il est possible d'obtenir sur l'ensemble de test considéré, sans toutefois préciser comment les obtenir dans la pratique (voir chapitre 4). Les experts utilisés seront supposés indépendants.

3.2 Fusion dure: les opérateurs ET et OU

On caractérisera ici le meilleur niveau de performance que l'on peut obtenir par l'utilisation des opérateurs ET et OU dans le cadre d'un schéma de fusion dure. Ces deux opérateurs seront également comparés entre eux et l'on mettra en évidence les plages de travail pour lesquelles un opérateur est préférable [52]. Cette section traite du cas particulier de deux experts. Si l'on envisage de fusionner trois experts ou plus, on veillera à travailler de façon diadique et itérative, en fusionnant le k^{ime} expert avec un expert fictif résultant de la fusion des $(k-1)$ experts restant.

3.2.1 Equations générales

Soit (fa_1, fr_1) et (fa_2, fr_2) deux points de fonctionnement pris sur les courbes caractéristiques (voir section 4.3, partie 1) de deux experts *indépendants*, référencés par les indices 1 et 2[1]. L'appartenance de ces points à la courbe caractéristique peut être explicitée de la façon suivante:

$$fr_i = CC_i(fa_i) \tag{3.1}$$

où $i = \{1, 2\}$, et $CC_i()$ désigne la fonction décroissante qui définit la courbe caractéristique de l'expert i. Comme fa_i et fr_i représentent les fractions des imposteurs acceptés ou des clients rejetés, nous avons:

$$0 \leq fa_i, fr_i \leq 1 \tag{3.2}$$

Les taux de FA et FR pour les fusions ET (le candidat est accepté si il est accepté par les 2 experts) et OU (le candidat est accepté dès que l'un des deux experts l'accepte) peuvent être exprimés de la façon suivante:

$$FA_{ou} = fa_1 + fa_2 - fa_1 fa_2 \tag{3.3}$$
$$FR_{ou} = fr_1 fr_2 \tag{3.4}$$
$$FA_{et} = fa_1 fa_2 \tag{3.5}$$

1. Par souci de clarté, nous adopterons dans le courant de ce chapitre des notations quelque peu différentes de celles introduites au chapitre 1.

$$FR_{et} = fr_1 + fr_2 - fr_1 fr_2 \tag{3.6}$$

L'établissement de ces équations est détaillé à l'annexe B et suppose des modalités indépendantes.

Dorénavant, les lettres majuscules désigneront les performances du système en aval de la fusion (performances du superviseur), tandis que les minuscules feront référence aux performances en amont (performances des experts).

Intuitivement, l'opérateur OU tend à minimiser le faux rejet tandis que l'opérateur ET, minimiser la fausse acceptation. En effet, à supposer que les deux experts fournissent des décisions contradictoires – l'un validant le candidat, l'autre le rejetant – une fusion OU acceptera l'individu et évitera ainsi le rejet d'un client éventuel. L'opérateur ET par contre, le rejettera et empêchera l'accès d'un possible imposteur. Cette constatation pouvait aussi être déduite du jeu d'équations (3.3) à (3.6) où l'on observe quels que soient les deux points de fonctionnement (fa_1, fr_1) et (fa_2, fr_2) choisis:

$$(3.4) \leq (3.6) \tag{3.7}$$
$$(3.5) \leq (3.3) \tag{3.8}$$

Ceci revient à placer sur un graphe FA/FR (figure 3.1, sous-figure de gauche), le point de fonctionnement obtenu par le superviseur ET, en haut et à gauche du point de fonctionnement relatif au OU. Ce comportement ne permet néanmoins pas de présager de la supériorité d'un opérateur sur l'autre puisque en modifiant les conditions opératoires et en passant du doublet $\{(fa_1, fr_1); (fa_2, fr_2)\}$ au nouveau doublet $\{(fa_1^*, fr_1^*); (fa_2^*, fr_2^*)\}$, un résultat contraire aux inégalités (3.7) et (3.8) peut être obtenu, à savoir:

$$(3.4)_* \geq (3.6) \tag{3.9}$$
$$(3.5)_* \geq (3.3) \tag{3.10}$$

où * se rapporte aux nouvelles conditions opératoires.

Ce cas est illustré de façon schématique sur la figure 3.1, où les croix et les cercles représentent les points opératoires obtenus par fusion ET et OU respectivement. En se limitant à la première sous-figure, on pourrait croire

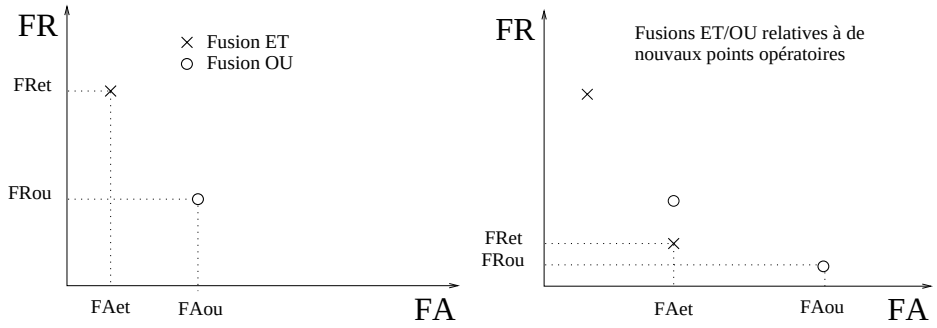


Figure 3.1 - *Comportement des opérateurs ET et OU*

que l'opérateur OU donne globalement de meilleurs résultats que le ET. Il existe néanmoins une autre combinaison de points de fonctionnement aboutissant aux performances illustrées à la seconde sous-figure, et pour laquelle l'opérateur ET réussit à faire mieux encore. Aussi, pour pouvoir déterminer sous quelles conditions un opérateur l'emporte sur l'autre, il nous faudra davantage développer les équations (3.3) à (3.6).

3.2.2 Cas particulier FA=0

Un premier résultat relativement simple peut être obtenu pour un superviseur dont on impose un point de fonctionnement à FA=0. Dans ce cas, les équations (3.3) et (3.5) doivent être égalées à zéro. Si l'on veut que (3.3) soit nulle, la condition (3.2) implique que $fa_1 = fa_2 = 0$, et l'on observe un taux de faux rejet:

$$FR_{ou|FA=0} \quad = \quad fr_1^0 fr_2^0 \tag{3.11}$$

où $fr_i^0 = CC_i(0)$. Pour que (3.5) soit nulle, il suffit qu'un seul des deux fa_i soit égal à zéro, soit $fa_1 = 0$ par exemple. On a alors:

$$FR_{et|FA=0} \quad = \quad fr_1^0 + fr_2 - fr_1^0 fr_2 \tag{3.12}$$

Suite aux inégalités (3.2), on peut montrer que quelque soit la valeur de fr_2, on a (3.11) $\leq$ (3.12). *Ainsi, pour FA=0, le minimum de FR est atteint par l'opérateur OU, avec comme résultat un FR exprimé par l'équation (3.11).*

En supposant des *courbes caractéristiques* $CC_i()$ *continues*, on peut étendre ce résultat et énoncer la propriété suivante: **pour un taux de fausse acceptation FA suffisamment faible, l'opérateur logique OU offre de meilleures performances de fusion que l'opérateur ET**.

Le domaine des FA "suffisamment faibles" pour lesquels l'opérateur OU offre de meilleures performances que le ET, peut fortement varier selon les valeurs de fr_i^0. Ainsi, pour des fr_i^0 proches de l'unité, ce domaine est tellement réduit que l'on ne peut plus considérer l'opérateur OU comme ayant une utilité pratique. Le cas d'un fr_i^0 proche de l'unité n'est heureusement que rarement rencontré puisque qu'il correspond à une modalité qui n'est pas capable de différencier un client d'un imposteur de façon fiable. Pour des valeurs de fr_i^0 usuelles (en deçà de 50% par exemple), la supériorité de l'opérateur OU s'étend sur une vaste étendue de FA faibles, comme l'illustre la figure 3.4 que nous commenterons par la suite.

3.2.3 Fusion OU à faible FA

Mieux encore, à partir de la connaissance des courbes caractéristiques $CC_i()$ des deux experts que l'on désire fusionner, nous pouvons ébaucher la courbe caractéristique du superviseur OU sous l'hypothèse de faibles FA. En approximant les courbes caractéristiques par leur développement au premier ordre autour de zéro et en considérant des valeurs de fausse acceptation suffisamment faibles, on peut approcher les équations (3.3) et (3.4) par:

$$\begin{aligned} FA_{ou} &\simeq fa_1 + fa_2 && (3.13)\\ FR_{ou} &\simeq (fr_1^0 - \alpha_1 fa_1)(fr_2^0 - \alpha_2 fa_2) \\ &\simeq fr_1^0 fr_2^0 - \alpha_1 fr_2^0 fa_1 - \alpha_2 fr_1^0 fa_2 && (3.14) \end{aligned}$$

où α_i représente la valeur absolue de la pente de la courbe caractéristique $CC_i()$ au point $(0, fr_i^0)$. Pour un FA_{ou} donné (qui fixe par conséquent la somme $fa_1 + fa_2$), le minimum de l'équation (3.14) est obtenu pour

$$fa_j = FA_{ou} \quad (3.15)$$
$$fa_{(1-j)} = 0 \quad (3.16)$$

où *l'indice j correspond à l'expert qui offre le produit $\alpha_j fr^0_{(1-j)}$ le plus élevé* et $(1-j)$ à l'autre expert. Nous pouvons alors réécrire l'équation (3.14) en tenant compte de ce résultat. Nous obtenons ainsi l'ensemble des points de fonctionnement optimaux dans le cadre d'une fusion OU:

$$\begin{aligned} FR_{ou} &\simeq fr^0_1 fr^0_2 - \alpha_j fr^0_{(1-j)} FA \\ &\simeq fr^0_{(1-j)}(fr_j - \alpha_j FA) \\ &\simeq fr^0_{(1-j)} CC_j(FA) \end{aligned} \quad (3.17)$$

Pour rappel, ce développement n'est valable qu'autour de zéro, c'est-à-dire uniquement pour de faibles taux de fausse acceptation. Ceci n'a que relativement peu d'importance, puisque ce sont de tels taux que l'on vise à obtenir dans la pratique et que s'en écarter n'a que peu de sens.

Résumons ce que l'on a obtenu jusqu'à présent. Nous avons montré qu'à faible taux de fausse acceptation, l'opérateur OU offre de meilleures performances que l'opérateur ET, et que celles-ci sont caractérisées par l'équation (3.17). Cette équation nous apprend en outre que la courbe caractéristique du superviseur OU suit à un facteur d'échelle $fr^0_{(1-j)}$ près, la courbe caractéristique de l'expert j qui offre le coefficient $\alpha_j fr^0_{(1-j)}$ le plus élevé.

La figure 3.2 représente les performances de deux experts hypothétiques (courbes caractéristiques en traits interrompus), ainsi que l'ensemble des points de fonctionnement du superviseur OU (points) obtenus en fusionnant toutes les combinaisons possibles de points opératoires $(fa_1, fr_1); (fa_2, fr_2)$ pris à intervalles réguliers le long de leurs courbes caractéristiques respectives. Cette figure représente également le résultat de l'approximation (3.17) (trait continu) qui, comme on le voit, fournit effectivement l'enveloppe des meilleures performances possibles du superviseur OU. Cette enveloppe a la même allure que la courbe caractéristique du premier expert (c'est en effet celui qui offre le coefficient $\alpha_j fr^0_{(1-j)}$ le plus élevé) au facteur d'échelle 0.5 près, soit la valeur de l'ordonnée à l'origine du deuxième expert.

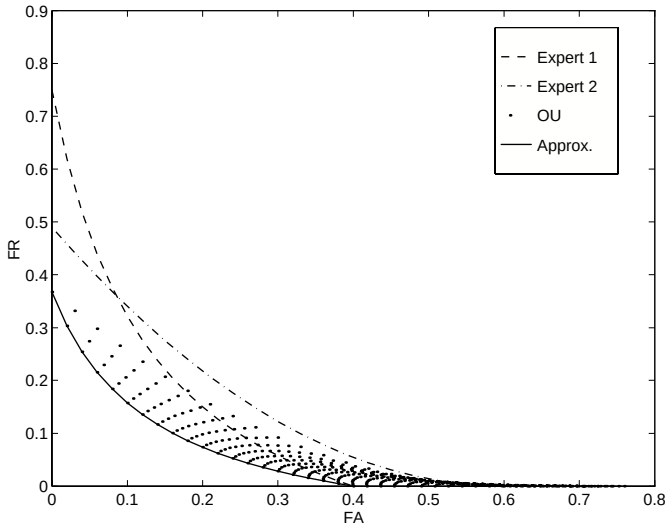


Figure 3.2 - *Superviseur OU : deux experts hypothétiques, l'ensemble des points de fonctionnement du superviseur et le résultat de l'estimation des meilleures performances possibles.*

3.2.4 Fusion ET à faible FR

Grâce à la symétrie des équations (3.3) à (3.6), des propriétés similaires à celles obtenues jusqu'à présent peuvent être transposées pour un superviseur basé sur l'opérateur ET, en prenant soin toutefois de remplacer le concept de fausse acceptation par celui de faux rejet. Ainsi, la propriété suivante peut être enoncée: **pour un taux de faux rejet FR suffisamment faible, l'opérateur logique ET offre de meilleures performances de fusion que l'opérateur OU**. Ces performances sont caractérisées par une relation identique à (3.17) valable pour des FR suffisamment faibles:

$$FA_{et} \simeq fa^{0}_{(1-j)} CC_j^{-1}(FR) \tag{3.18}$$

où $CC_j^{-1}()$ désigne la fonction réciproque de $CC_j()$.

La figure 3.3 reprend les deux mêmes experts que ceux de la figure 3.2, mais les combine en utilisant cette fois l'opérateur ET. A nouveau, on voit comment l'équation (3.18) approche fidèlement les meilleures performances

qu'il est possible d'obtenir, pour autant que l'on respecte l'hypothèse sous laquelle cette équation a été établie, à savoir de faibles taux de FR.

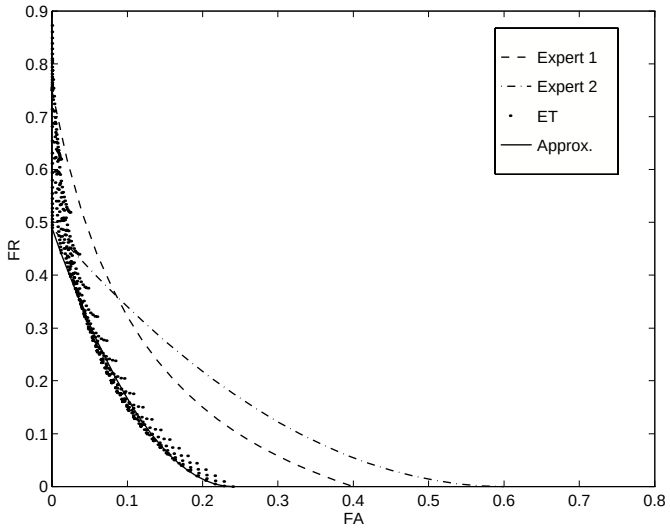


Figure 3.3 - *Superviseur ET : deux experts hypothétiques, l'ensemble des points de fonctionnement du superviseur et le résultat de l'estimation des meilleures performances possibles.*

3.2.5 Superviseur hybride OU/ET

La figure 3.4 illustre comment les résultats issus des deux sous-sections précédentes se complètent idéalement. Cette figure compare entre elles les deux courbes résultant des approximations (3.17) et (3.18) précédemment illustrées en traits continus dans les figures 3.2 et 3.2. On y remarque les meilleures performances du OU à faible taux de fausse acceptation et celles du ET à faible taux de faux rejet. Mieux encore, on voit comment les approximations (3.17) et (3.18), qui supposaient respectivement de faibles FA et FR, sont justement valables là ou leurs opérateurs respectifs se révèlent être les plus utiles. Ces approximations peuvent donc être utilisées pour déterminer les meilleures performances que peut offrir un système ayant recours à une fusion de type dure (soit l'enveloppe inférieure des deux courbes représentées en tirets et en continu sur la figure 3.4), combinant de façon optimale les opérateurs ET et OU.

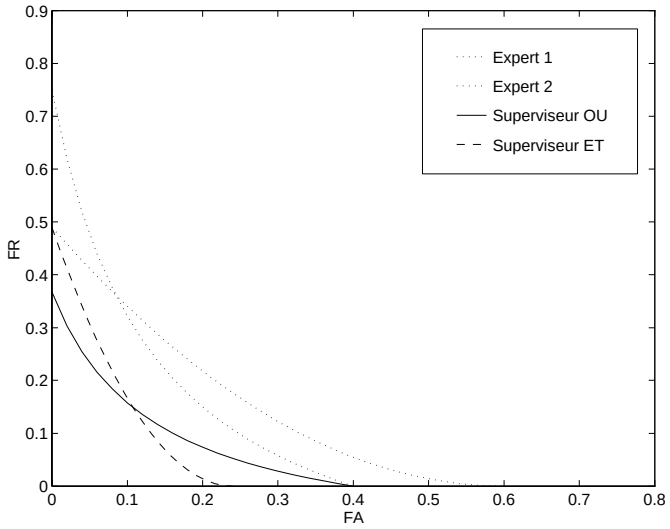


Figure 3.4 - *Comparaison des superviseurs ET et OU*

3.3 Fusion douce: combinaison linéaire de scores

Etudions à présent un des superviseurs doux les plus couramment utilisés. Il se base sur une combinaison linéaire des scores en provenance des différents experts[2].

Alors que dans un schéma de fusion dure, la seule connaissance des courbes caractéristiques des experts que l'on fusionne suffisait pour pouvoir caractériser les performances du superviseur, il devient nécessaire à présent de descendre jusqu'aux scores fournis par les différents experts, puisque c'est à ce niveau là que s'opérera la fusion proprement dite.

Deux types de distributions ont été considérés, à savoir les distributions uniformes et normales (ou gaussiennes). Une distribution uniforme répartit de façon équiprobable les scores entre deux bornes données. On veillera à choisir des plages plus élevées pour la classe clients que pour celle des imposteurs, tout en se préservant une zone de recouvrement afin d'éviter le cas trivial où dès le départ on disposerait d'un expert parfait (taux d'erreur

2. Les méthodes non linéaires débouchent généralement sur de meilleurs résultats mais sont beaucoup plus sensibles au manque de données d'apprentissage. Nous y reviendrons dans le courant de la section 4.5.3.

nul). Une distribution normale suppose une répartition de scores gaussienne et caractérise la classe clients par une moyenne plus élevée que celle des imposteurs. Une répartition gaussienne des scores sous-entend un domaine de distribution infini, avec comme résultat, l'impossibilité de rejeter tous les imposteurs sans devoir rejeter l'ensemble des clients également.

Selon une règle de fusion linéaire, un candidat est accepté si la somme pondérée des scores relatifs à différents experts dépasse un seuil d'acceptation fixé, ce qui donne dans le cas de deux experts:

$$\lambda z^{(1)} + (1 - \lambda) z^{(2)} > k \tag{3.19}$$

où λ représente un facteur compris entre [0,1] permettant de pondérer davantage l'avis d'un expert par rapport à l'autre.

Des performances typiques de fusions linéaires dans le cadre de distributions uniformes et normales sont données aux figures 3.5 et 3.6. Les courbes relatives au superviseur linéaire (traits continus) fournissent les meilleures performances possibles et sont obtenues en effectuant pour chaque point de fonctionnement (FA, FR) une recherche exhaustive sur le paramètre λ. Ces courbes sont comparées aux meilleures performances que l'on aurait pu obtenir par l'utilisation d'un schéma de fusion dure de type ET ou OU. Les résultats sont sensiblement différents d'un graphe à l'autre. Sous l'hypothèse de distributions uniformes, la fusion dure donne de meilleurs résultats que la fusion linéaire pour de faibles taux de FA ou FR, tandis que cette dernière est préférable en milieu de dynamique. Sous l'hypothèse de distributions gaussiennes, un superviseur linéaire surpasse un superviseur hybride ET/OU quelque soit le point de fonctionnement envisagé. Ceci peut être expliqué par le fait que sous l'hypothèse gaussienne, il n'est pas possible d'avoir moins de 100% de faux rejet lorsque FA=0, ou de façon symétrique, moins de 100% de fausse acceptation lorsque FR=0. Comme nous avons pu le constater en commentant les équations (3.3) à (3.6), ceci correspond à une situation où les opérateurs ET et OU ne sont d'aucune utilité.

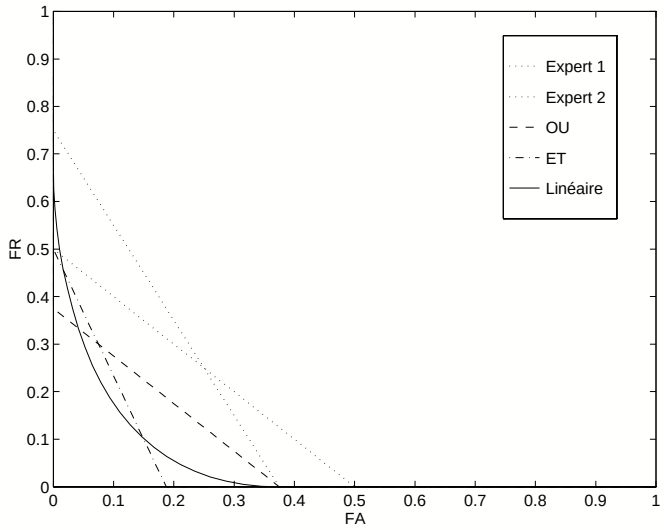


Figure 3.5 - *Comparaison entre fusions dures et douces sous l'hypothèse de distributions de scores uniformes*

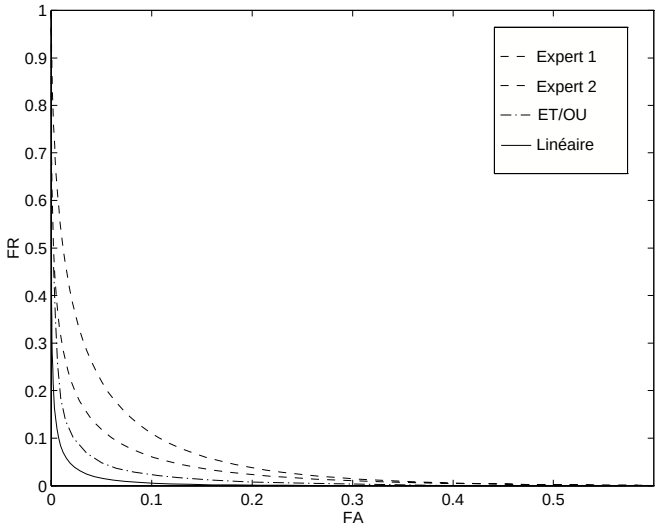


Figure 3.6 - *Comparaison entre fusions dures et douces sous l'hypothèse de distributions de scores gaussiennes*

3.4 Commentaires

Afin de pouvoir comparer les superviseurs de type dur ou doux et de déterminer quelle fusion est à même d'offrir les meilleures performances, il est nécessaire de disposer des répartitions statistiques des scores clients et imposteurs relatifs aux experts que l'on désire fusionner. Par l'étude réalisée aux sections 3.2 et 3.3, nous avons pu mettre en évidence que les mérites d'un système de fusion par rapport à l'autre dépendent non seulement du type de ces répartitions mais également de la zone de fonctionnement que l'on désire atteindre. D'un point de vue théorique, aucun des superviseurs étudiés jusqu'à présent n'a donc pu émerger comme étant la réponse unique au problème général posé par la fusion de données.

Soulignons le danger qui pourrait se présenter à vouloir déterminer de façon théorique le gain qui résulterait de la fusion de superviseurs dont les répartitions de scores auraient été approchées par des modèles théoriques comme ceux utilisés au point précédent. Ces approximations peuvent parfois se révéler suffisantes pour biaiser la validité pratique du résultat théorique, comme l'ont mis en évidence certaines simulations effectuées dans le cadre de ce travail. La prudence s'impose donc.

Puisque la modélisation du comportement d'un expert pratique par une distribution statistique risque, par ses approximations, de porter préjudice aux conclusions que l'on pourrait obtenir, on ne peut mieux conseiller que d'utiliser des données réelles acquises durant une session d'apprentissage et d'en faire directement usage afin de déterminer le type de superviseur qui convient pour les besoins de l'application. Le recours à des données réelles pour effectuer le choix d'un superviseur ne doit pas être considéré comme un inconvénient. Ces données, qui semblent à présent être nécessaires, l'étaient déjà précédemment dans le cas où l'on envisageait une approche purement théorique, puisqu'il faut pouvoir à un moment ou un autre mettre en correspondance le modèle utilisé avec les observations récoltées sur le terrain (trouver les valeurs des moyennes et variances d'une modélisation gaussienne par exemple). La section suivante évalue les performances des différents types de fusion envisagés ici, sur les *données réelles* fournies par les experts spécifiquement développés dans le cadre de ce travail.

3.5 Résultats expérimentaux

Les deux types de superviseurs ont été testés sur nos experts profil et frontal afin d'en sélectionner le meilleur en vue de son éventuelle implémentation au sein du système final. Chaque expert fait ici usage des seuils d'acceptation individuels tels que définis dans la première partie de ce travail. Les performances sont calculées sur l'ensemble des 10656 tests générés par le protocole de test expert décrit au chapitre 3 de la partie 1. Les courbes illustrées à la figure 3.7 font à nouveau référence aux meilleures performances qu'il est possible d'obtenir sur cet ensemble de test, pour un superviseur donné. Comme on le voit, dans le cas particulier de la base de données M2VTS et dans le cas de la fusion de nos deux experts profil et frontal, un superviseur de type dur fournit les meilleurs résultats. En outre, on remarque que l'opérateur OU surpasse l'opérateur ET sur pratiquement la totalité des points de fonctionnement. Ce comportement aurait pu être prédit à partir de l'analyse théorique du point 3.2, comme l'illustre la figure 3.8. Cette figure représente deux courbes caractéristiques grossièrement similaires à celles qui caractérisent nos experts ainsi que les performances des deux superviseurs ET et OU obtenues par les équations (3.17) et (3.18). L'opérateur OU y offre bien de meilleures performances sur près de la totalité du domaine de fonctionnement.

Les performances du meilleur superviseur, à savoir le superviseur OU dans notre cas, sont résumées dans le tableau 3.1. Elles débouchent sur une amélioration sensible des performances par comparaison à celles offertes par les experts pris individuellement (TEE de 7% dans le meilleur des cas).

	Seuillage	TEE	TS	$TFR^{1\%}$
Superviseur OU	Individuel	2%	96.5%	3%

Tableau 3.1 - *Performances du superviseur OU lors de la fusion des experts profil et frontal*

En incluant l'expert vocal aux deux experts précédents, les meilleures performances sont alors obtenues par le *superviseur linéaire* avec un taux d'erreur totale inférieur à 0.05%, soit un TS supérieur à 99.95%.

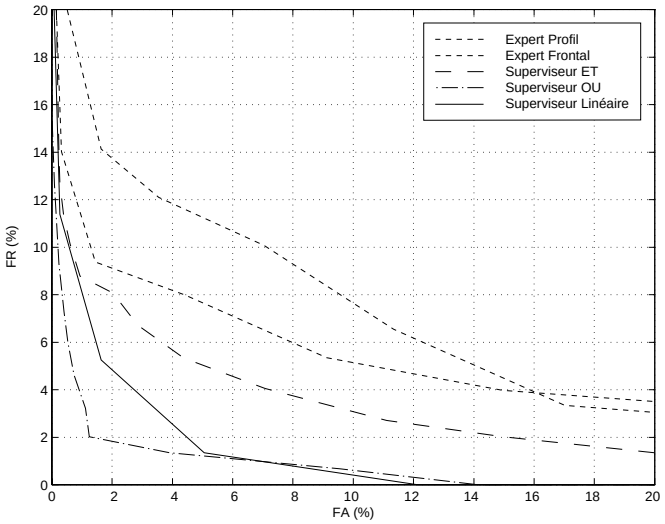


Figure 3.7 - *Comparaison entre fusions dures et douces sur des données expérimentales : fusion entre les experts profil et frontal*

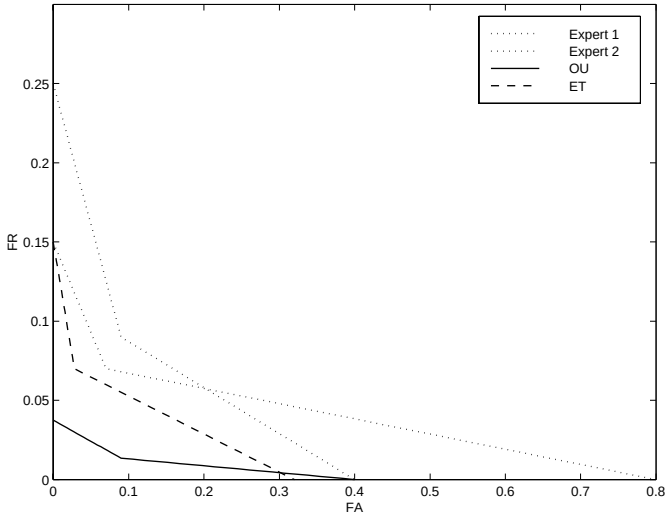


Figure 3.8 - *Comparaison entre opérateurs ET et OU sur des courbes caractéristiques approximativement similaires à celles des experts profil et frontal*

Chapitre 4

Analyse a priori (test sur des scores inconnus)

4.1 Contexte

Dans la section 3.5, nous avons comparé différents superviseurs sur base de leurs meilleures performances observées sur un ensemble de test pratique. Cet ensemble de test unique était par ailleurs identique à celui utilisé pour caractériser les performances de nos experts. Plusieurs remarques doivent être faites quant à la validité et la portée des résultats obtenus au chapitre précédent.

D'une part, même si il nous a été possible de déterminer des performances optimales, nous ne savons toujours pas *comment* les obtenir a priori. Ces performances sont le fruit d'une recherche exhaustive sur l'ensemble des combinaisons de seuils possibles (fusion dure) ou des combinaisons linéaires (fusion douce) et ce n'est qu'une fois toutes ces combinaisons envisagées qu'ont pu être isolées celles qui donnent les meilleurs résultats.

D'autre part – et c'est important – cette recherche s'effectuait sur un unique ensemble de test, court-circuitant ainsi toute phase d'apprentissage préliminaire du superviseur. Cela revient à optimiser un système sous l'hypothèse d'une connaissance préalable des clients et des imposteurs qui se présenteront lors du test. Les résultats ainsi obtenus sont qualifiés de résultats *a posteriori* puisqu'ils ont été obtenus *en connaissant le statut client/imposteur*

de chaque sujet, ainsi que leurs scores respectifs. Une telle procédure ne correspond évidemment pas à la réalité des choses, et nous verrons ici comment adopter une procédure de test plus réaliste qualifiée d'*a priori.*

Les résultats du chapitre précédent nous sont tout de même d'une utilité appréciable, à savoir qu'ils fournissent une borne supérieure aux performances qu'il nous est permis d'obtenir sur l'ensemble de test.

Enfin, la théorie précédemment établie supposait l'indépendance des experts utilisés. Nous nous départirons de cette hypothèse dans le présent chapitre.

4.2 Protocole de test superviseur

Si l'on veut pouvoir caractériser les performances réelles dont bénéficierait un superviseur placé dans des conditions opératoires pratiques, il est nécessaire de travailler avec des ensembles d'entraînement et de test distincts. Durant l'entraînement (ou l'apprentissage), le superviseur dispose d'un ensemble de scores clients et imposteurs qui lui permettent de calibrer les valeurs de ses paramètres internes (seuils d'acceptation, coefficients des combinaisons linéaires, ...) et d'optimiser un critère fonction des TFA et TFR souhaités. Une fois l'entraînement terminé, les paramètres du superviseur sont fixés et celui-ci est alors placé dans des conditions opératoires réelles. Dans notre cas, ces conditions seront simulées par l'intermédiaire d'un ensemble de test distinct de celui utilisé lors de l'apprentissage. En phase de test, le superviseur est alors amené à rencontrer de nouveaux clients et des imposteurs véritables (tout à fait inconnus du système).

Le choix des ensembles d'apprentissage et de test pour le superviseur doit être judicieux. Tout d'abord, on vient de le voir, ces deux ensembles doivent être disjoints et ne peuvent en aucun cas utiliser les mêmes données si l'on veut rester fidèle à la réalité. Ensuite, il faut éviter que l'entraînement du superviseur ne se fasse sur les mêmes données que celles utilisées lors de l'entraînement des experts, soit les images de référence dans le cas des experts profil et frontal. En effet, sur de telles données, les experts ne pourront produire que d'excellents résultats. Le superviseur serait alors entraîné sur des experts qu'il estimerait erronément être bons et fiables. Pour éviter un tel biais, nous devrons obligatoirement entraîner le superviseur sur les données obtenues en phase de test de nos experts. En outre, ces données de-

vront également être utilisées lors du test du superviseur puisque ici encore, l'utilisation des données d'apprentissage relatives aux experts n'est pas admissible. Ceci implique que l'ensemble de test des experts devra à la fois servir lors de l'apprentissage et lors du test du superviseur. Ces deux phases devant faire usage de données distinctes, nous devons nécessairement scinder l'ensemble de test des experts en deux sous-ensembles disjoints, appelés les sous-ensemble d'*entraînement* et *de test du superviseur* respectivement.

Pour rappel, les données de *test des experts* sont constituées de la prise de vues mise à l'écart, lors du choix des images de référence pour chaque client de la base de données M2VTS. Parmi les 37 personnes qui composent cette prise de vues, on compte 36 clients et un imposteur véritable [1] (voir partie 1, section 3.4). Tant la prise de vues de test que l'imposteur véritable sont des paramètres du protocole de test expert et peuvent être fixés librement.

Afin de générer les sous-ensembles d'apprentissage et de test du superviseur, il a été procédé cette fois comme suit:

- les individus de la base de données M2VTS ont été divisés en deux groupes de taille (quasi) équivalente, soit les individus de 1 à 18 (initiales allant de BP à JR) et de 19 à 37 (JT à XM). Ces groupes sont référencés par la suite sous les vocables Groupe 1 et Groupe 2 respectivement;
- pour chaque groupe, le protocole de test expert est utilisé pour générer des accès clients et imposteurs véritables. Ainsi pour le Groupe 1, nous disposerons des tests clients suivants:

  $$BP_{can} \leftrightarrow BP_{ref}$$
  $$BS_{can} \leftrightarrow BS_{ref}$$
  $$...$$
  $$JR_{can} \leftrightarrow JR_{ref}$$

  où les indices "can" et "ref" désignent respectivement les accès candidats provenant de la prise de vues mise à l'écart et les modèles de référence générés à partir des trois autres prises de vues. Les tests imposteurs au sein de ce même groupe seront constitués des mises en correspondance de:

  $$BP_{imp} \leftrightarrow BS_{ref},\ BP_{imp} \leftrightarrow CC_{ref},\ ...\ ,\ BP_{imp} \leftrightarrow JR_{ref}$$

1. Par opposition aux imposteurs simulés, obtenus en faisant passer un *client* pour un autre.

$BS_{imp} \leftrightarrow BP_{ref}$, $BS_{imp} \leftrightarrow CC_{ref}$, ...
...
$JR_{imp} \leftrightarrow BP_{ref}$, $JR_{imp} \leftrightarrow BS_{ref}$, ...

en prenant soin bien entendu d'utiliser dans ce cas les imposteurs mis à l'écart durant le protocole de test expert, comme candidats (indices "imp").

Nous procéderons de façon équivalente au sein du Groupe 2. Notons le cloisonnement strict entre les deux groupes (une personne appartenant à un groupe donné n'ira jamais imposter un individu du groupe adverse). En prenant soin d'effectuer la rotation entre prises de vues, ce protocole superviseur donne naissance à 72 tests clients et 1224 tests imposteurs dans le premier groupe, 76 tests clients et 1368 tests imposteurs dans le second.

Ces deux groupes serviront tour à tour d'ensemble d'entraînement et de test superviseur, étant entendu que si l'on entraîne sur l'un, on teste sur l'autre. En moyennant les résultats obtenus lors des deux sessions de test, on obtiendra une bonne estimation des performances *pratiques* que l'on peut attendre du superviseur.

Le protocole de test superviseur est illustré à la figure 4.1. On y fait en outre le lien avec le protocole expert défini dans la première partie de ce travail. Il est important d'y remarquer l'utilisation commune par les experts et le superviseur, de la personne mise à l'écart, comme imposteur.

La caractéristique principale d'un tel protocole superviseur est d'introduire une séparation totale entre l'ensemble d'entraînement et l'ensemble de test: tant les clients que les imposteurs dont il est fait usage dans les deux ensembles sont différents[2]. *Si cela représente un avantage certain au niveau des imposteurs, il est néanmoins regrettable que les clients utilisés lors des tests soient différents de ceux disponibles lors de l'entraînement.* Ceci supprime toute utilisation possible de seuils individuels[3], technique qui s'était avérée particulièrement efficace dans la première partie de ce travail. Cette pénalité est un mal nécessaire si l'on veut pouvoir tester le superviseur de façon correcte en ne faisant usage que des quatre prises de vues dont nous disposons pour tester à la fois les experts et le superviseur. Dans de telles

2. Tel est le cas lorsqu'un superviseur est optimisé en usine, sur un premier ensemble de personnes, puis testé "sur site" avec des clients et des imposteurs différents.

3. Ces seuils acquis sur un ensemble de clients donnés, ne peuvent bien entendu s'appliquer qu'à cet ensemble.

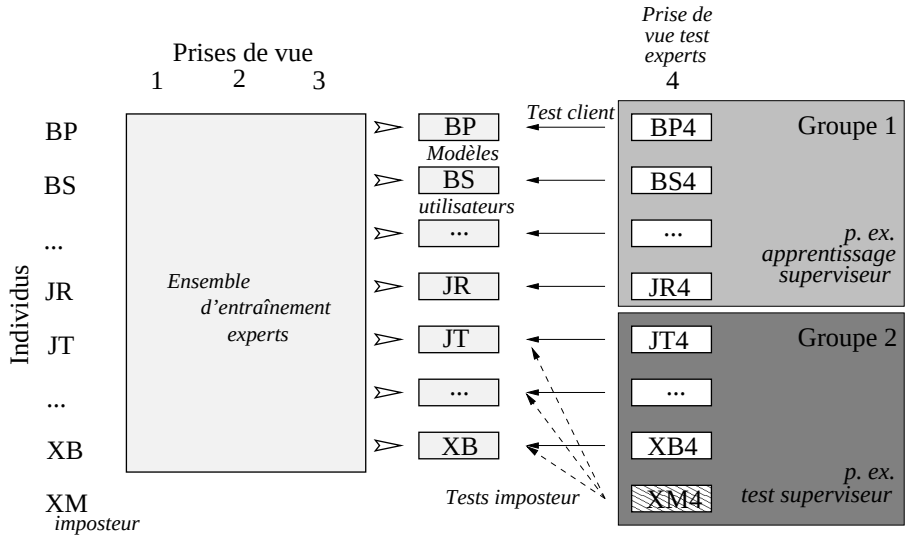


Figure 4.1 - *Protocole de test superviseur: illustration d'un test client dans le Groupe 1 et d'un test imposteur au sein du Groupe 2 (dans le cas d'un test imposteur, il est essentiel d'utiliser l'imposteur défini par le protocole de test expert).*

conditions, nous pouvons supposer que les performances obtenues dans la suite de ce travail sont inférieures à celles que l'on aurait pu réellement obtenir si nous avions disposé de plus de prises de vues lors de l'apprentissage de chaque client.

4.3 Performance des experts selon le nouveau protocole de test

Ayant défini un protocole de test différent de celui utilisé dans la première partie de ce travail, il est bon de caractériser à nouveau le niveau de performance atteint par chaque expert pris individuellement. Ces performances nous seront utiles par la suite pour pouvoir juger du gain effectif apporté par la fusion de nos experts.

Le tableau 4.1 présente les FA et FR relatifs aux différents experts utilisés

dans le cadre de ce travail et obtenus en optimisant différents critères sur l'ensemble d'apprentissage. Ces critères sont au nombre de quatre, à savoir:

- minimisation de la FR à FA nulle
- minimisation de la FA à FR nulle
- minimisation du TEE
- maximisation du TS

Le tableau 4.1 dissocie les deux ensembles de test (groupe 1 et groupe 2) et nous fournit pour chacun d'eux:

- les FA et FR optimaux, c'est-à-dire les meilleures performances qu'il est possible d'obtenir au sein du groupe considéré (colonnes "Apprentissage");
- les FA et FR réellement obtenus, c'est-à-dire les performances obtenues en phase de test, associées aux seuils qui maximisent le niveau de performance lors de l'entraînement dans le groupe dual (colonnes "Test").

Ce tableau doit être interprété de la façon suivante (illustration de la première ligne): pour l'expert profil, et selon un critère qui minimise le taux de FR à FA nulle, un apprentissage sur le groupe 1 fournit au mieux $FA = 0\%$ et $FR = 23.6\%$. En appliquant les conditions opératoires optimales issues de cet apprentissage sur le groupe 2, nous obtenons alors des performances de $FA = 0.3\%$ et $FR = 38.2\%$ (phase de test). En intervertissant les rôles des groupes 1 et 2, nous obtenons les performances suivantes: $FA = 0\%$ et $FR = 51.3\%$ lors de l'apprentissage (groupe 2) et $FA = 0\%$ et $FR = 38.9\%$ lors du test (groupe 1). On peut également interpréter les colonnes "apprentissage" et "test" d'un même groupe, comme évaluant respectivement les performances a posteriori et a priori de l'expert. Ces deux colonnes permettent donc de quantifier l'écart par rapport à l'optimum des performances qui résultent d'une phase d'apprentissage indépendante de celle du test.

Le tableau 4.2 présente ces performances de façon plus synthétique, en moyennant les résultats obtenus dans chaque groupe.

Expert	Critère	Groupe 1				Groupe 2			
		Apprent.		Test		Apprent.		Test	
		FA	FR	FA	FR	FA	FR	FA	FR
Profil	FA=0	0	23.6	0	38.9	0	51.3	0.3	38.2
	FR=0	51.4	0	64.4	0	62.9	0	49.3	1.3
	TEE	TEE 7.0		7.0	6.9	TEE 6.6		5.9	9.2
	TS	TS 87.8		TS 83.3		TS 86.3		TS 83.0	
Frontal	FA=0	0	40.3	0	45.8	0	21.1	1.2	18.4
	FR=0	71.2	0	61.2	1.4	87.5	0	92.0	0
	TEE	TEE 8.2		0.9	27.8	TEE 8.3		28.8	5.3
	TS	TS 83.7		TS 66.5		TS 85.3		TS 65.9	
Vocal	FA=0	0	2.8	0	4.2	0	1.3	0.2	1.3
	FR=0	2.9	0	1.6	1.4	2.8	0	3.7	0
	TEE	TEE 1.5		1.6	1.3	TEE 1.5		2.8	0
	TS	TS 98.5		TS 95.8		TS 98.7		TS 98.1	

Tableau 4.1 - *Performances des experts individuels évaluées selon le nouveau protocole de test superviseur (valeurs exprimées en %)*

Expert	Critère	Apprent.		Test	
		FA	FR	FA	FR
Profil	FA=0	0	37.5	0.2	38.6
	FR=0	57.2	0	56.7	0.7
	TEE	TEE 6.8		6.5	8.1
	TS	TS 87.1		TS 83.2	
Frontal	FA=0	0	30.7	0.6	32.1
	FR=0	79.4	0	76.6	0.7
	TEE	TEE 8.3		14.9	16.6
	TS	TS 84.5		TS 66.2	
Vocal	FA=0	0	2.1	0.1	2.8
	FR=0	2.9	0	2.7	0.7
	TEE	TEE 1.5		2.2	0.7
	TS	TS 98.6		TS 97.0	

Tableau 4.2 - *Valeurs moyennes des performances des experts individuels évaluées selon le nouveau protocole de test superviseur (valeurs exprimées en %)*

Ces tableaux montrent, à nouveau, l'excellent comportement de l'expert vocal, malgré l'utilisation d'un seuillage global. Le tableau 4.2 met bien en évidence les écarts que l'on peut observer entre les performances optimales et celles réellement obtenues. En général, les dégradations constatées restent acceptables, hormis le cas de l'expert frontal pour lequel l'ensemble d'entraînement semble ne pas être représentatif de l'ensemble de test.

4.4 Superviseur exhaustif

La première famille de superviseurs étudiée met en œuvre une technique d'entraînement simple et intuitive. Sur base des données d'apprentissage, les paramètres internes du superviseur sont optimisés afin d'en maximiser les performances sur cet ensemble d'entraînement. La recherche des meilleurs paramètres du superviseur se fera ici de façon exhaustive, c'est-à-dire en envisageant toutes les combinaisons possibles et en ne retenant que celle qui optimise un critère de performance donné. Une fois ces paramètres fixés, le superviseur est évalué sur l'ensemble de test. Nous qualifierons un tel superviseur de *superviseur exhaustif*.

4.4.1 Mise en œuvre

Pour les superviseurs durs ET et OU, toutes les combinaisons de seuils k_i possibles sont testées sur l'ensemble d'entraînement. La combinaison qui offre les meilleures performances est alors sélectionnée. En supposant N experts distincts, nous pouvons formuler la procédure d'apprentissage de la manière suivante:

$$k_i^* \;\mid\; (z^{(1)} \geq k_1^*) \oslash (z^{(2)} \geq k_2^*) \oslash \cdots \oslash (z^{(N)} \geq k_N^*) \iff minf(FA, FR) \tag{4.1}$$

où k_i^* représentent les seuils optimaux, $z^{(i)}$ les scores fournis par chacun des experts, $i = 1..N$, $\oslash$ l'opérateur ET ou OU et $f(FA, FR)$ le critère à optimiser.

Dans le cas d'un superviseur doux basé sur une combinaison linéaire des scores, l'optimisation portera sur les coefficients de la combinaison linéaire

et le seuil global k:

$$k^*, \alpha_1^* \ \cdots \ \alpha_{n-1}^* \ | \\ \alpha_1^* z^{(1)} + \alpha_2^* z^{(2)} + \cdots + \alpha_{N-1}^* z^{(N-1)} + (1 - \alpha_1^* - \alpha_2^* - \cdots - \alpha_{N-1}^*) z^{(N)} \geq k^* \\ \iff minf(FA, FR) \tag{4.2}$$

Comme les scores $z^{(i)}$ sont normalisés entre [0,1], la recherche exhaustive pour chacun des paramètres k_i^* ou k s'effectuera sur le domaine [0,1] également, discrétisé dans la pratique en 21 incréments de pas de 0.05. Il en va de même pour les coefficients α_i^* qui prennent en compte toutes les combinaisons possibles allant de l'expert seul ($\alpha_j = 1$, $alpha_i = 0$, $\forall i \neq j$) à une pondération identique de tous les experts ($\alpha_i = 1/N$, $\forall i$). A nouveau, 21 pas ont été utilisés pour discrétiser la dynamique de chaque coefficient.

4.4.2 Résultats expérimentaux

Le tableau 4.3 reprend de façon détaillée les résultats issus de la fusion a priori des experts profil et frontal, par les superviseurs exhaustifs décrits ci-dessus. Il s'interprète de façon identique au tableau 4.1, à la seule différence près que pour chaque critère, deux résultats sont fournis, l'un relatif à la fusion dure (seul le meilleur opérateur y est repris, il est référencé entre parenthèses), l'autre à la fusion douce (linéaire).

Le tableau 4.4 fournit les performances obtenues lors de la fusion de l'ensemble des experts dont nous disposons, à savoir les experts profil, frontal et vocal.

4.4.3 Commentaires

Le tableau 4.5 récapitule les performances issues des tableaux précédents et moyenne les résultats obtenus sur les deux groupes d'individus. A partir de ce tableau, plusieurs constatations peuvent être faites. Commençons par le cas de la fusion des deux experts profil et frontal (colonne EP+EF):

- à fausse acceptation nulle (1ère ligne), contrainte que l'on désire souvent respecter en pratique, les deux types de fusion offrent des performances similaires, avec une préférence toutefois pour le superviseur dur de type OU. Curieusement, celui-ci se révèle être légèrement

EP + EF									
Critère	Superviseur	Groupe 1				Groupe 2			
		Apprent.		Test		Apprent.		Test	
		FA	FR	FA	FR	FA	FR	FA	FR
FA=0	Dur (OU)	0	9.7	0	18.1	0	11.8	1.8	9.2
	Doux	0	9.7	0.1	19.4	0	7.9	3.7	14.5
FR=0	Dur (ET)	33.1	0	35.5	1.4	48.3	0	41.5	1.3
	Doux	4.8	0	15.9	0	35.2	0	15.2	4.0
TEE	Dur (OU)	TEE 2.8		0.4	15.3	TEE 4.7		13.4	5.3
	Doux	TEE 1.8		2.9	1.4	TEE 6.6		5.4	6.6
TS	Dur (OU)	TS 95.3		TS 92.7		TS 92.5		TS 75.9	
	Doux	TS 96.6		TS 87.4		TS 92.9		TS 88.2	

Tableau 4.3 - *Performances des différents superviseurs exhaustifs évaluées selon le nouveau protocole de test superviseur (valeurs exprimées en %). Fusion des experts profil et frontal.*

EP + EF + EV									
Critère	Superviseur	Groupe 1				Groupe 2			
		Apprent.		Test		Apprent.		Test	
		FA	FR	FA	FR	FA	FR	FA	FR
FA=0	Dur (OU)	0	0	0	1.4	0	0	2.4	0
	Doux	0	0	0	1.4	0	1.3	0.5	3.9
FR=0	Dur (ET)	0.9	0	0.4	2.8	1.5	0	1.7	1.3
	Doux	0	0	0	1.4	0.2	0	0.5	3.9
TEE	Dur (OU)	TEE 0		0	1.4	TEE 0		2.4	0
	Doux	TEE 0		0	1.4	TEE 0.2		0.5	3.9
TS	Dur (OU)	TS 100		TS 97.6		TS 100		TS 98.6	
	Doux	TS 100		TS 98.6		TS 99.8		TS 95.6	

Tableau 4.4 - *Performances des différents superviseurs exhaustifs évaluées selon le nouveau protocole de test superviseur (valeurs exprimées en %). Fusion des experts profil, frontal et vocal.*

Critère	Superviseur	EP+EF				EP+EF+EV			
		Apprent.		Test		Apprent.		Test	
		FA	FR	FA	FR	FA	FR	FA	FR
FA=0	Dur (OU)	0	10.8	0.9	13.7	0	0	1.2	0.7
	Doux	0	8.8	1.9	17.0	0	0.7	0.3	2.7
FR=0	Dur (ET)	40.7	0	38.5	1.4	1.2	0	1.1	2.1
	Doux	10.4	0	25.6	2.0	0.1	0	0.3	2.7
TEE	Dur (OU)	TEE 3.8		6.9	10.3	TEE 0		1.2	0.7
	Doux	TEE 4.2		4.2	4.0	TEE 0.1		0.3	2.7
TS	Dur (OU)	TS 93.9		TS 84.3		TS 100		TS 98.1	
	Doux	TS 94.8		TS 87.8		TS 99.9		TS 97.1	

Tableau 4.5 - *Récapitulatif des performances moyennes obtenues pour la fusion exhaustive de deux ou trois experts (valeurs exprimées en %)*

moins bon sur l'ensemble d'apprentissage, mais néanmoins plus robuste puisqu'il débouche sur de meilleurs résultats lors du test proprement dit;

– tant les superviseurs durs que doux ne donnent que de piètres résultats lorsqu'on travaille à taux de faux rejet nul (2ème ligne). Alors que l'on pense obtenir de relativement bonnes performances avec le superviseur linéaire, celles-ci se dégradent considérablement sur l'ensemble de test. Le superviseur dur donne de moins bons résultats encore;

– dans toutes les situations intermédiaires (FA et FR non nuls mais faibles, tels ceux atteints en utilisant les critères TEE et TS), la fusion douce semble représenter la solution de choix.

Si, à partir de ce tableau, il nous fallait choisir un superviseur offrant le meilleur compromis entre faux rejet et fausse acceptation, nous opterions probablement pour le superviseur linéaire optimisé sur base du critère TEE. Celui-ci offre, tant sur les ensembles d'apprentissage que de test, un taux d'égale erreur moyen de 4%, ce qui représente une amélioration par un facteur deux, environ, des performances que l'on aurait pu obtenir en utilisant

un expert unique[4].

En ce qui concerne l'usage de trois experts, toutes les techniques se révèlent être équivalentes à FA=0 et FR=0. Par contre, pour les critères TEE et TS, c'est le superviseur de type dur, et plus particulièrement l'opérateur OU, qui l'emporte. Le meilleur compromis est sans doute atteint pour ce même opérateur et le critère TEE. On observe alors un taux d'égale erreur moyen de 1% environ. En dépit de nos attentes, ces résultats ne sont que légèrement meilleurs que ceux observés en présence de l'expert vocal seul. Pour rappel, celui-ci était capable d'offrir un TEE de 1.5% environ ($FA = 2.2\%$ et $FR = 0.7\%$ plus précisément).

Remarquons également que les valeurs des performances obtenues durant la phase d'apprentissage se situent souvent bien au-delà de ce que l'on obtient en réalité sur l'ensemble de test, quelque soit le nombre d'experts envisagé. Ceci donne à penser que la technique d'apprentissage exhaustive utilisée ici, n'est pas assez robuste et que des répartitions de scores clients et imposteurs légèrement différentes entre ensembles d'apprentissage et de test suffisent à perturber notre superviseur.

4.5 Superviseur statistique

Les résultats quelque peu décevants du superviseur exhaustif peuvent être expliqués par un phénomène appelé *surapprentissage*: un superviseur peut tellement bien prendre en compte les caractéristiques particulières de l'ensemble d'entraînement qu'il devient incapable de pouvoir traiter tout autre ensemble n'offrant pas strictement les mêmes caractéristiques, comme l'ensemble de test par exemple. C'est en réalité ce qui se passe avec notre superviseur exhaustif: durant l'apprentissage, la combinaison optimale des paramètres du superviseur est choisie après avoir essayé toutes les combinaisons possibles sur l'ensemble d'entraînement. Cette combinaison optimale se révèle être conditionnée par les clients ou les imposteurs ayant un comportement particulier[5], c'est-à-dire les individus les moins représentatifs de la population dont ils sont issus. Ces clients ou imposteurs particuliers ayant disparu dans l'ensemble de test (et peut-être ayant été remplacés

4. L'expert profil, soit le meilleur des deux experts utilisés offrait un TEE moyen de plus de 7% sur l'ensemble de test (voir section 4.3)

5. Un client qui serait accepté par tous les experts, à l'exception de l'un d'eux qui le rejette catégoriquement, par exemple.

par d'autres individus tout aussi non représentatifs mais aux particularités différentes), le superviseur perdra automatiquement l'efficacité acquise lors de l'entraînement. **Il est donc dangereux d'avoir à faire à un superviseur trop bon lors de l'apprentissage, de crainte que celui-ci ne prenne en compte des caractéristiques d'une population qui ne se retrouveront plus lors du test**. On voit ainsi apparaître un compromis entre *les bonnes performances obtenues lors de l'entraînement* et la *robustesse du superviseur face à de nouvelles conditions opératoires*. Ceci motive l'usage d'une modélisation globale des populations clients et imposteurs lors de l'apprentissage et le recours aux statistiques pour caractériser aux mieux les propriétés "macroscopiques" de ces ensembles. Tel est le cas du superviseur présenté ci-après.

Notons que le surapprentissage n'est pas une fatalité en soi et qu'il peut être combattu en utilisant un ou plusieurs ensembles de validation en plus de l'ensemble d'apprentissage. Un nombre élevé de données est néanmoins requis pour pouvoir générer tous ces ensembles. Ces données feront malheureusement défaut dans le cadre du protocole de test superviseur défini sur la (relativement petite) base de données M2VTS.

4.5.1 Superviseur de Fisher

Le superviseur *statistique* introduit ici, se base sur les travaux de Fischer[6] [24] et fait usage d'une frontière de décision linéaire pour séparer deux populations données, à savoir les clients et les imposteurs dans notre cas. Dans le contexte particulier de l'authentification de personnes, [19] utilise lui aussi une règle de décision linéaire, mais l'optimise en se fixant un critère dérivé de la théorie de bayes [4]. D'autres classificateurs sont étudiés par [8] ainsi que [66] dans le contexte général de la discrimination de populations multiples.

6. Ronald Aylmer Fisher, né à Londres (Angleterre) en 1890, mort à Adélaide (Australie) en 1962. Diplomé en astronomie de l'université de Cambridge en 1912, il appliqua la théorie des erreurs aux observations astronomiques puis s'orienta par la suite vers la biologie en 1919. Ses travaux l'ont amené à s'intéresser aux problèmes de nature statistique et à développer de nombreux concepts probabilistes bien connus aujourd'hui comme le maximum de vraisemblance, les tests d'hypothèses, l'analyse de la variance (ANOVA) et l'étude de distributions statistiques adaptées aux échantillons de taille réduite. Fisher est considéré comme l'un des fondateurs des statistiques modernes en raison de ses nombreuses contributions importantes dans ce domaine.

Envisageons à présent la règle de décision développée par Fisher. Elle se base sur le rapport de vraisemblance (1.26) repris ci-dessous:

$$\frac{T(z|c)}{T(z|i)} > k \tag{4.3}$$

Pour rappel, k représente un seuil d'acceptation dont la valeur dépend du compromis TFA/TFR que l'on veut obtenir. Dans le problème qui nous préoccupe, $T(z|c)$ et $T(z|i)$ sont inconnus et doivent être estimés à partir des données d'apprentissage. Une hypothèse courante consiste à approcher les distributions réelles par des distributions normales à p variables $N_p(\mu_A, \Sigma)$, où $A = \{c, i\}$ représente la classe d'individus, μ_A le vecteur des scores moyens et Σ la matrice de covariance entre experts. En un premier temps, on supposera la matrice Σ indépendante de la classe d'individus. Sous de telles hypothèses, les fonctions de densité de probabilité s'écrivent sous la forme:

$$f_A(z) = (2\pi)^{-p/2}\ |\Sigma|^{-1/2}\ exp\left\{-\frac{1}{2}(z-\mu_A)'\ \Sigma^{-1}\ (z-\mu_A)\right\} \tag{4.4}$$

Les paramètres μ_c, μ_i et Σ sont inconnus, mais peuvent être estimés à partir des données d'apprentissage, soit x les n_c données relatives aux accès clients et y, les n_i données relatives aux accès imposteurs (simulés). On a :

$$\widehat{\mu}_c = \sum_{q=1}^{n_x} x_q/n_c \tag{4.5}$$

$$\widehat{\mu}_i = \sum_{q=1}^{n_u} y_q/n_i \tag{4.6}$$

$$\widehat{\Sigma}_c = \sum_{q=1}^{n_x} (x_q - \widehat{\mu}_c)(x_q - \widehat{\mu}_c)'/(n_c - 1) \tag{4.7}$$

$$\widehat{\Sigma}_i = \sum_{q=1}^{n_y} (y_q - \widehat{\mu}_i)(y_q - \widehat{\mu}_i)'/(n_i - 1) \tag{4.8}$$

$$\widehat{\Sigma} = [(n_c - 1)\widehat{\Sigma}_c + (n_i - 1)\widehat{\Sigma}_i]/(n_c + n_i - 2) \tag{4.9}$$

Notons que l'on tient compte ici, par l'intermédiaire de Σ, de la dépendance qui peut exister entre experts.

4.5.2 Mise en œuvre

En combinant les équations (4.3) à (4.9), on peut réécrire $\widehat{f}_c(z)/\widehat{f}_i(z) \geq k$ sous la forme de $D_L(z) \geq ln(k) = k^*$ où

$$D_L(z) = (z - \frac{1}{2}(\widehat{\mu}_c + \widehat{\mu}_i))' \ \widehat{\Sigma}^{-1} \ (\widehat{\mu}_c - \widehat{\mu}_i) \tag{4.10}$$

Fisher fut le premier à utiliser cette fonction à des fins de classification. Comme $D_L(z)$ est linéaire en z, elle fut communément appelée *fonction linéaire discriminante*[7]. Ainsi, la procédure à suivre pour vérifier l'identité d'un candidat, consiste à calculer $\widehat{\mu}_c$, $\widehat{\mu}_i$ et $\widehat{\Sigma}$ à partir des données d'entraînement (ce qui est effectué une fois pour toutes), puis $D_L(z)$ et comparer celui-ci au seuil k^* donné. Si $D_L(z) \geq k^*$, le candidat est accepté comme client.

4.5.3 Superviseur quadratique

Dans le cas de distributions de scores clients et imposteurs ne satisfaisant pas l'hypothèse d'une covariance Σ unique, la règle de décision (4.10) peut se réécrire sous la forme $D_Q(z) \geq 2k^*$ où

$$\begin{aligned} D_Q(z) \ &= \ (z - \widehat{\mu}_i)' \ \widehat{\Sigma}_i^{-1} \ (z - \widehat{\mu}_i) \\ &\quad - \ (z - \widehat{\mu}_c)' \ \widehat{\Sigma}_c^{-1} \ (z - \widehat{\mu}_c) \\ &\quad + \ ln(|\widehat{\Sigma}_i|/|\widehat{\Sigma}_c|) \end{aligned} \tag{4.11}$$

$D_Q(z)$ est appelée *fonction quadratique discriminante*[8]. Cette règle permet d'obtenir de meilleurs résultats que ceux obtenus dans le cadre d'une fusion linéaire, mais requiert un ensemble d'apprentissage suffisamment étendu pour pouvoir évaluer Σ_c et Σ_i avec précision. Dans le cas du protocole de test superviseur adopté à la section 4.2, le nombre réduit de données d'apprentissage (relatives aux tests clients en particulier) ne permet pas d'évaluer ces covariances de façon individuelle, et nous contraint à l'hypothèse d'une covariance Σ unique.

7. ou LDF, *Linear Discriminant Function*, en anglais.
8. ou QDF, *Quadratic Discriminant Function*, en anglais.

4.5.4 Résultats expérimentaux

De manière identique à celle dont les superviseurs précédents avaient été caractérisés, le tableau 4.6 reprend les performances obtenues par le *superviseur statistique linéaire*, dans le cadre de la fusion de deux ou trois experts. La colonne "Apprentissage" correspond aux résultats obtenus sur l'ensemble d'entraînement en utilisant la règle de décision (4.10) et des paramètres $\widehat{\mu}$ et $\widehat{\Sigma}$ estimés sur ce même ensemble. Il ne s'agit donc plus d'un optimum au sens de la section 4.3. La colonne "test" applique la règle (4.10) sur le groupe d'individus qui s'y réfère, mais fait usage des paramètres $\widehat{\mu}$ et $\widehat{\Sigma}$ préalablement estimés *sur le groupe dual*[9].

Remarquons les performances remarquables obtenues pour le groupe 1 dans le cadre de la fusion de trois experts (TEE=0%).

Experts	Critère	Groupe 1				Groupe 2			
		Apprent.		Test		Apprent.		Test	
		FA	FR	FA	FR	FA	FR	FA	FR
EP+EF	FA=0	0	11.1	0	13.9	0	9.2	0.2	9.2
	FR=0	16.8	0	26.6	0	41.2	0	29.0	1.3
	TEE	TEE 3.0		3.4	2.8	TEE 6.3		4.5	6.6
	TS	TS 94.1		TS 91.2		TS 92.7		TS 88.9	
EP+EF+EV	FA=0	0	0	0	0	0	1.3	0.4	1.3
	FR=0	0	0	0	0	0.1	0	0	2.6
	TEE	TEE 0		0	0	TEE 0.05		0.4	1.3
	TS	TS 100		TS 100		TS 99.95		TS 98.3	

Tableau 4.6 - *Performances du superviseur statistique évaluées selon le nouveau protocole de test superviseur (valeurs exprimées en %). Fusion des experts profil et frontal, ainsi que profil, frontal et vocal.*

9. Théoriquement, les performances mentionnées dans la colonne "Apprentissage" du superviseur exhaustif (tableaux 4.3 et 4.4) devraient être meilleures que celles obtenues lors de l'apprentissage du superviseur statistique (tableau 4.6), puisqu'elles sont sensées fournir les meilleurs résultats possibles. En pratique, la recherche "exhaustive" des paramètres optimaux du superviseur exhaustif s'effectue sur un ensemble de valeurs discrètes (voir section 4.4.1). De ce fait, il est possible de passer à coté de la solution optimale. Tel est le cas pour quelques-unes des valeurs de performance mentionnées dans les tableaux 4.3, 4.4 et 4.6. Ceci laisse présager l'excellent comportement du superviseur statistique.

4.5.5 Commentaires

Le tableau 4.7 résume les données précédentes et met en évidence la robustesse de la méthode proposée: les performances que laissent présager les simulations en environnement d'apprentissage sont nettement plus fidèles à ce qui est réellement obtenu lors du test, que dans le cas des superviseurs exhaustifs. Aussi, le niveau de performance obtenu en phase de test est à chaque fois meilleur que ceux obtenus précédemment (voir tableau 4.5) et particulièrement dans le cas d'un superviseur combinant l'entièreté des experts disponibles. Nous obtenons alors un taux de réussite de 99.2%. *Le taux d'erreur que nous aurions observé en faisant usage du meilleur expert, l'expert vocal, se trouve ainsi réduit par un facteur trois.*

Experts	Critère	Apprent.		Test	
		FA	FR	FA	FR
EP+EF	FA=0	0	10.2	0.1	11.6
	FR=0	29.0	0	27.8	0.7
	TEE	TEE 4.7		4.0	4.7
	TS	TS 93.4		TS 90.1	
EP+EF+EV	FA=0	0	0.7	0.2	0.7
	FR=0	0.1	0	0	1.3
	TEE	TEE 0		0.2	0.7
	TS	TS 100		TS 99.2	

Tableau 4.7 - *Récapitulatif des performances moyennes obtenues pour la fusion statistique de deux ou trois experts (valeurs exprimées en %)*

4.6 Incertitude sur les mesures de performance[10]

Attardons-nous enfin sur la pertinence des résultats chiffrés, présentés tout au long des sections précédentes. Ceux-ci sont en réalité des *estimations* de performances "réelles" inconnues, et dépendent en grande partie de la nature des individus qui composent les groupes d'apprentissage et de test précédemment définis. On peut raisonnablement penser que ces valeurs seraient toutes autres pour des individus différents.

10. Je tiens à remercier Benoît Maison pour l'aide apportée dans le cadre de cette section.

Dans cette section, nous veillerons à estimer un intervalle de confiance pour chacune des mesures de performance établies dans le cadre de la section 4.5[11]. La méthode choisie pour caractériser ces intervalles est numérique[12] et connue sous le nom de *technique du bootstrap* [22]. En voici sa description.

Soit y l'ensemble des caractéristiques biométriques relatives aux individus utilisés lors des accès clients ou imposteurs sur les ensembles d'entraînement et de test. En distinguant les deux groupes d'individus précédemment définis, nous pouvons expliciter y de la façon suivante:

$$y = \{v_1, v_2, ..., v_m, w_1, w_2, ..., w_n\} \tag{4.12}$$

où v_i se rapporte aux caractéristiques biométriques d'un individu du groupe 1 et w_j à celles d'un individu du groupe 2 ($m = 18$, $n = 19$). Nous supposerons les caractéristiques biométriques v et w réparties selon une loi de distribution F identique, mais inconnue:

$$v_*, w_* \sim F \tag{4.13}$$

A partir de y, nous avons pu estimer les performances d'un superviseur donné selon différents critères, par exemple en fournissant les TFA et TFR sur l'ensemble de test résultant d'une minimisation du TEE sur l'ensemble d'entraînement. Notons $\widehat{\Theta}(y)$ le paramètre estimé, soit TFA ou TFR dans l'exemple précédent. Dans cette section, nous nous intéressons au calcul de la variance de $\widehat{\Theta}(y)$, ou plutot de son écart-type que nous noterons $\sigma(F)$, car il dépend de la distribution F (entre-autres). La technique du *bootstrap* décrite ci-dessous permet d'estimer cet écart-type, c'est-à-dire de calculer un $\widehat{\sigma}(F)$. Dans [22], il est démontré que:

$$\widehat{\sigma}(F) = \sigma(\widehat{F}) \tag{4.14}$$

où $\widehat{F}$ représente une estimation discrète de la distribution inconnue F, qui associe à chaque échantillon v_* et w_* observé un poids identique $\frac{1}{\#echantillons}$.

11. Nous restreindrons notre analyse à cette seule section vu le temps de calcul élevé requis par la méthode proposée, particulièrement dans le cadre de l'implémentation d'un superviseur exhaustif.

12. Les densités de probabilité requises pour envisager une approche analytique sont inconnues et l'on ne désire pas faire ici, l'approximation classique de répartitions de scores gaussiennes.

Venons-en à la technique du *bootstrap*. Soit

$$y^* = \{v_1^*, v_2^*, ..., v_m^*, w_1^*, w_2^*, ..., w_n^*\} \tag{4.15}$$

un nouvel ensemble de caractéristiques biométriques obtenu en réorganisant le vecteur y de la façon suivante: dans chaque groupe d'individus v et w pris séparément, on effectue un tirage *avec remise* d'un nouvel échantillon de taille identique à celle de l'échantillon original. La procédure du *bootstrap* procède alors en trois étapes:

1. en utilisant un générateur de nombres aléatoires, on tire indépendamment un nombre élevé d'ensembles y^*, soit $y^*(1), y^*(2), ..., y^*(B)$;
2. sur chaque ensemble $y^*(b)$, $b = 1..B$, on évalue le paramètre Θ dont on cherche à caractériser l'écart-type, soit $\widehat{\Theta}^*(b) = \widehat{\Theta}(y^*(b))$;
3. on calcule l'écart-type relatif à la répartition des valeurs $\widehat{\Theta}^*(b)$, soit $\widehat{\sigma}_B$.

Pour $B \to \infty$, nous pouvons montrer que [22]:

$$\widehat{\sigma}_B = \widehat{\sigma}(F) = \sigma(\widehat{F}) \tag{4.16}$$

En résumé, la technique du bootstrap permet de calculer de façon numérique la variance d'un paramètre estimé sur des données dont la répartition est inconnue mais qui peut être approchée par une densité de probabilité discrète associant à chaque donnée de départ un poids identique.

[22] préconise des valeurs de B entre 50 et 200 pour la plupart des applications. Par prudence, nous avons opté pour $B = 1000$.

La technique du bootstrap fut appliquée aux résultats présentés dans la section 4.5, c'est-à-dire aux performances offertes par le superviseur statistique linéaire. On raffine ainsi les résultats expérimentaux mentionnés au tableau 4.7 en fournissant pour chacune des valeurs qui s'y trouve, un intervalle de confiance. Ces intervalles sont exprimés au tableau 4.8 sous la forme $\mu \pm \sigma$ où μ désigne la moyenne du paramètre estimé et σ son écart-type, calculés sur les résultats des 1000 tirages du bootstrap. Pour information, l'intervalle $\mu \pm \sigma$ offre une confiance de 70% (cas d'une loi

normale), les intervalles $\mu \pm 2\sigma$ et $\mu \pm 3\sigma$, une confiance de 95% et 99.998% respectivement [13].

Experts	Critère	Apprentissage		Test	
		FA	FR	FA	FR
EP+EF	FA=0	0 ± 0	10.37 ± 1.03	0.12 ± 0.11	10.83 ± 0.97
	FR=0	29.27 ± 1.07	0 ± 0	28.34 ± 1.16	0.65 ± 0.15
	TEE	*TEE 4.55 ± 0.33*		*4.31 ± 0.43*	*4.57 ± 0.40*
	TS	*TS 93.42 ± 0.47*		*TS 89.91 ± 0.69*	
EP+EF +EV	FA=0	0 ± 0	0.48 ± 0.32	0.26 ± 0.15	0.41 ± 0.34
	FR=0	0.07 ± 0.07	0 ± 0	0 ± 0.01	1.31 ± 0.22
	TEE	*TEE 0.04 ± 0.05*		*0.07 ± 0.07*	*0.78 ± 0.31*
	TS	*TS 99.93 ± 0.07*		*TS 99.15 ± 0.30*	

Tableau 4.8 - *Intervalles de confiance pour l'évaluation des performances du superviseur statistique, calculées selon la méthode du bootstrap (intervalles de confiance à 70%, soit $\mu \pm \sigma$ – valeurs exprimées en %).*

Les figures 4.2 à 4.5 représentent les distributions réellement obtenues pour quelques-unes des performances exprimées sous forme d'intervalle au tableau 4.8 (performances relatives aux critères TEE et TS, voir lignes reprises en italique dans le tableau 4.8).

Comme nous le montre le tableau 4.8, les intervalles de confiance calculés par la méthode du bootstrap sont tous relativement étroits (écart-type de 0.3% en moyenne). Ce résultat permet de valider les mesures de performance du superviseur statistique présentées à la section 4.5, mais aussi de confirmer la supériorité du superviseur statistique sur le superviseur exhaustif, l'incertitude de mesure sur les performances du superviseur statistique étant inférieure à l'écart de performance noté entre les deux types de superviseurs.

13. Cette confiance exprime la probabilité que la valeur réelle recherchée soit effectivement dans l'intervalle considéré.

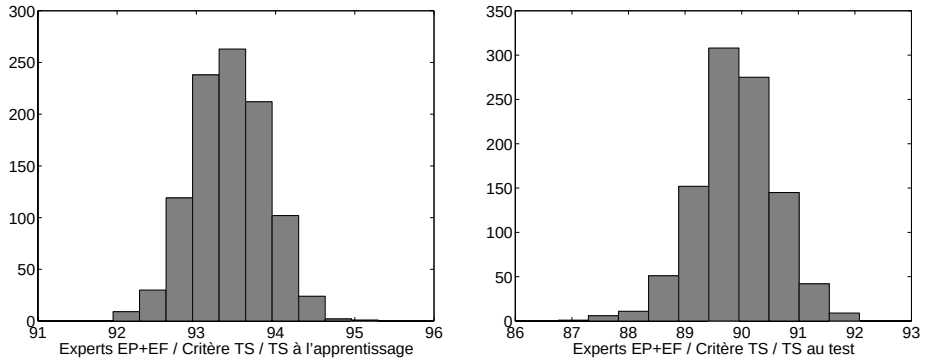


Figure 4.2 - *Détail des distributions de scores engendrés par la technique du bootstrap. Cas des experts EP+EF, critère du TS (valeurs exprimées en %).*

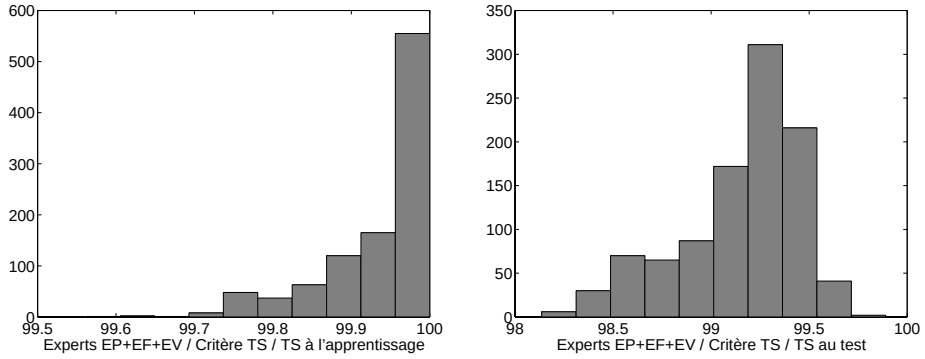


Figure 4.3 - *Détail des distributions de scores engendrés par la technique du bootstrap. Cas des experts EP+EF+EV, critère du TS (valeurs exprimées en %).*

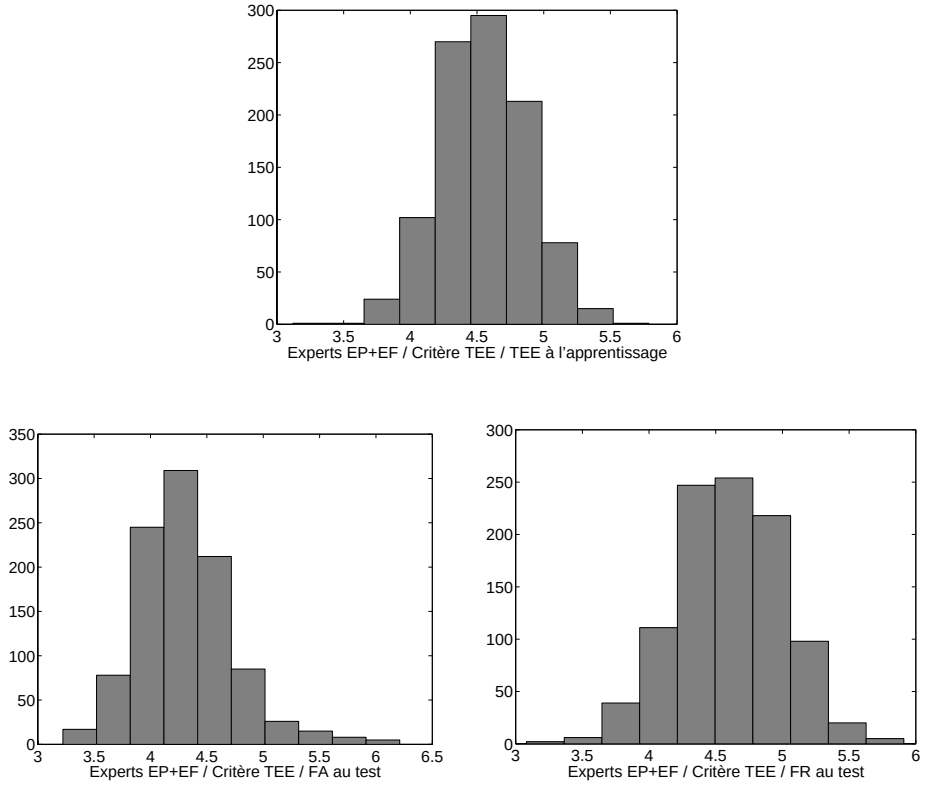


Figure 4.4 - *Détail des distributions de scores engendrés par la technique du bootstrap. Cas des experts EP+EF, critère du TEE (valeurs exprimées en %).*

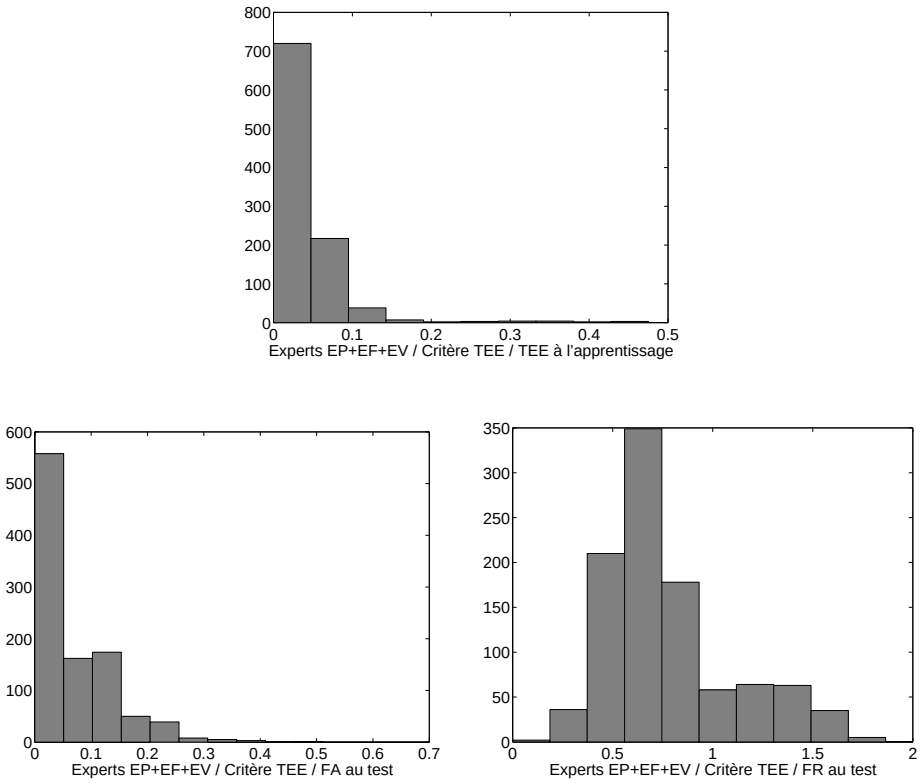


Figure 4.5 - *Détail des distributions de scores engendrés par la technique du bootstrap. Cas des experts EP+EF+EV, critère du TEE (valeurs exprimées en %).*

Chapitre 5

Conclusion de la seconde partie

Cette seconde partie fut consacrée à l'étude du superviseur, soit le module chargé de collecter l'information fournie par les différents experts et de prendre une décision finale quant à l'acceptation ou le rejet de l'individu que l'on authentifie.

Un premier chapitre formalisait le problème de la fusion de données dans le contexte d'une authentification d'identité. Une solution générique sous forme d'un rapport de vraisemblance y fut apportée. Cette solution se révéla optimale tant au sens des critères de type Neyman-Pearson, qu'en suivant une approche bayesienne. Les superviseurs étudiés dans le cadre de cette seconde partie furent ensuite introduits sur base du formalisme adopté, nous permettant ainsi de mieux cerner leur spécificité.

Au cours du deuxième chapitre, nous avons mis en évidence quelques propriétés fondamentales liées à la fusion de données. Tout d'abord, et contrairement à une idée largement répandue, les conditions requises à l'obtention d'un gain de fusion maximal ne sont pas nécessairement liées à l'indépendance des experts traités. Nous avons pu montrer l'avantage qui découlait de l'utilisation d'experts corrélés négativement. Par la suite, nous nous sommes intéressés à l'influence que pouvait avoir l'hypothèse d'indépendance – hypothèse fréquemment utilisée dans la littérature – sur la validité des résultats obtenus en présence d'experts dépendants. Cette influence existe, mais semble être négligeable pour des taux de corrélation de l'ordre

de ceux qui caractérisent des modalités que l'on devine être relativement indépendantes, comme par exemple la modalité profil avec la modalité vue de face ou les modalités images avec la modalité parole.

Dans le troisième chapitre, nous avons caractérisé de façon analytique, une borne aux meilleures performances qu'il est possible d'obtenir en fusionnant des experts définis par leurs courbes caractéristiques. Cette borne fut obtenue pour une fusion de type dure, la fusion douce requérant une information plus détaillée que celle fournie par les courbes caractéristiques utilisées. Ensuite, nous avons comparé de façon théorique les niveaux de performance issus des fusions dures et douces respectivement. Aucun type de superviseur n'a pu émerger comme solution idéale au problème posé en fusion de données, leurs mérites respectifs s'étant avérés trop sensibles à la nature des données que l'on souhaite fusionner, ainsi que du point opératoire que l'on désire obtenir. Faute de n'avoir pu isoler un schéma de fusion idéal, l'ensemble des opérateurs de fusion étudiés dans le cadre de ce travail (ET, OU et linéaire) furent appliqués aux experts développés dans la première partie de ce travail. Concernant les experts profil et frontal, ce fut l'opérateur dur OU qui apporta le meilleur gain de fusion avec un taux de succès de 96.5%. Combinant ces deux experts à l'expert vocal, ce taux dépassa 99.95% pour une fusion de type linéaire cette fois. L'analyse théorique menée dans le cadre de ce chapitre supposait l'indépendance des experts utilisés et le recours à un ensemble de test unique connu du superviseur (analyse *a posteriori*).

Enfin, le quatrième chapitre fut quant à lui dédié à l'étude du superviseur placé dans des conditions opératoires réelles, soit optimisé et testé sur des ensembles de données distincts (analyse *a priori*). Ces ensembles furent judicieusement choisis parmi les données de la base de données M2VTS, c'est-à-dire en ayant pris soin qu'aucun individu n'apparaisse simultanément dans les ensembles d'apprentissage et de test. Ces conditions, assez restrictives en soi, correspondent au système dont les paramètres auraient été optimisés en usine sur un premier ensemble de personnes, puis testé sur site avec des clients et des imposteurs différents. En général, l'apprentissage est également réalisé sur site afin de tenir compte des caractéristiques propres à l'ensemble des clients définitifs. Cela n'a pu se faire ici faute d'une quantité suffisante de données (bien plus de quatre prises de vues par client sont alors nécessaires si l'on veut pouvoir entraîner tant les experts que le superviseur sur des données disjointes). Deux types de superviseurs furent étudiés dans le cadre de ce chapitre. Le premier se basait sur la recherche

exhaustive des meilleurs paramètres à affecter au superviseur, le second sur une modélisation statistique de celui-ci (fonction linéaire discriminante de Fisher). La recherche exhaustive se révéla être fort sensible au phénomène de surapprentissage. On préféra ainsi utiliser un superviseur statistique, plus robuste et plus performant. Pour ce dernier, on nota un taux de succès de 90.1% pour la combinaison des experts profil et frontal, et ce en phase de test, 99.2% en leur adjoignant l'expert vocal. Contrairement au chapitre précédent, la modélisation statistique utilisée ici tenait compte de la dépendance éventuelle entre experts. Un intervalle de confiance fut calculé pour chaque estimation des performances du superviseur statistique. Ces intervalles relativement réduits ont pu confirmer l'excellent comportement du superviseur statistique et sa supériorité face au superviseur exhaustif.

Conclusions générales et développements futurs

Conclusions générales

Le présent travail avait pour but de développer un système d'authentification d'identité qui soit facile et peu coûteux d'implémentation, et d'en étudier objectivement les performances. Différentes modalités furent combinées afin de bénéficier d'une robustesse accrue face à la variabilité des caractéristiques physiques d'une personne. Notre choix s'est porté sur deux modalités images, l'une travaillant sur le visage vu de profil et l'autre sur le visage de face, ainsi qu'une modalité liée à la parole. Les informations relatives à ces modalités, aussi appelées *experts* en raison de leur expertise développée dans un domaine biométrique particulier, sont collectées au *superviseur*, qui se charge de prendre la décision ultime quant au rejet ou à l'acceptation de la personne qui s'identifie.

Ce travail fut divisé en deux parties. La première partie traitait de la conception et de l'étude des experts profil (partie 1, chapitre 1), frontal (partie 1, chapitre 2) et vocal (partie 1, chapitre 5). Tout au long de celle-ci, une attention toute particulière fut portée sur différents points, tels

- l'usage d'algorithmes rapides caractérisés par un compromis aussi élevé que possible entre la simplicité de leur mise en œuvre et le niveau de performance offert (partie 1, sections 1.2, 1.3.4, 1.4, 2.2 et 2.3);
- la résolution de problèmes pratiques, rendant possible l'automatisation des méthodes proposées dans le cadre d'une implémentation concrète. Il s'agissait, entre autres, de problèmes tels que l'extraction de contours sur des images faiblement contrastées (partie 1, section 1.4) ou la localisation automatique des yeux à partir d'une image du visage vu de face (partie 1, section 2.3);
- l'élaboration d'un protocole de test rigoureux afin de mesurer le ni-

veau de performance réel de nos différentes modalités (partie 1, section 3.4);

– l'utilisation d'une base de données de visages/voix offrant une variabilité des caractéristiques biométriques traitées aussi proche possible de celle que nous serions amenés à traiter en pratique (partie 1, chapitre 3).

La seconde partie fut dédiée au superviseur, le module responsable de la fusion des données en provenance de nos différents experts. Cette section fut également mise à profit pour approfondir certains aspects encore peu investigués en fusion de données, à savoir que

– contrairement à une idée largement répandue, les conditions requises à l'obtention d'un gain de fusion maximal ne sont pas nécessairement liées à l'indépendance des experts traités (partie 2, section 2.2). Nous avons pu montrer l'avantage qui découlait de l'utilisation d'experts corrélés négativement (partie 2, section 2.4);

– l'hypothèse d'indépendance entre experts est une hypothèse fréquemment utilisée dans la littérature. Nous avons voulu mettre en évidence son influence sur la validité des résultats obtenus en présence d'experts dépendants. Pour des experts caractérisés par une même fiabilité et dont la dépendance peut être approchée par une relation linéaire, cette influence peut être considérée comme nulle pour des taux de corrélation de l'ordre de ceux qui caractérisent des modalités que l'on devine être indépendantes, et encore négligeable pour des taux de corrélation relativement élevés (partie 2, section 2.6).

Concernant l'étude du superviseur proprement dit, les contributions originales de ce travail se situent à divers niveaux:

– une solution générique au problème de l'authentification multimodale d'identité a pu être établie. Elle s'exprime sous forme d'un rapport de vraisemblance et se révéla optimale au sens des critères de Neyman-Peason et de Bayes, mais également pour d'autres critères usuels, comme la maximisation du taux de succès ou la minimisation du taux d'égale erreur (partie 2, sections 1.4 et 1.5). Le formalisme mathématique utilisé pour arriver à ce résultat nous a en outre permis de

caractériser les spécificités propres à chaque superviseur traité dans le cadre de cette seconde partie.

- une borne aux meilleures performances qu'il est possible d'obtenir en fusionnant des experts définis par leur courbe caractéristique, a pu être obtenue de façon analytique pour un schéma de fusion de type dur. L'excellente complémentarité des opérateurs logiques ET et OU y a aussi été démontrée (partie 2, section 3.2).

- différents types de superviseurs représentatifs des schémas de fusion dure et douce furent comparés sur le plan théorique. Malheureusement, aucun d'eux n'a pu émerger comme solution générale au problème posé en fusion de données, leurs mérites relatifs s'étant révélés trop sensibles à la nature des données que l'on fusionne, ainsi que du point opératoire désiré (partie 2, section 3.4);

- une comparaison approfondie des superviseurs durs et doux fut effectuée sur le plan pratique, en faisant usage des données en provenance des experts précédemment étudiés. Cette comparaison fut menée dans le cadre de deux contextes différents mais complémentaires. Une étude *a posteriori* déboucha sur les meilleures performances qu'il était possible d'obtenir en faisant usage d'un opérateur de fusion donné et en supposant connu l'ensemble des scores clients/imposteurs utilisés pour caractériser ces performances (partie 2, chapitre 3). L'étude *a priori* plaça ensuite ces différents superviseurs dans des conditions opératoires plus réelles, dissociant les données d'apprentissage utilisées pour optimiser un superviseur donné, des données de test utilisées pour caractériser ses performances[1] (partie 2, chapitre 4). Dans ce dernier cas, nous avons pu montrer un net avantage à vouloir modéliser de façon statistique le superviseur que l'on désire optimiser (partie 2, section 4.5);

- dans tous les cas d'étude, le gain appréciable apporté par la fusion de données a clairement été établi.

Terminons sur quelques résultats quantitatifs et rappelons ici les performances de notre système final. Celui-ci résulte de la fusion des experts profil, frontal et vocal et fait usage d'un superviseur de type linéaire basé sur la modélisation statistique des données d'apprentissage. Il offre un taux

1. Ces deux ensembles de données avaient été confondus lors de l'étude a posteriori.

de faux rejet de 0.7% pour un taux de fausse acceptation de 0.2%, et un taux de succès de 99.2% (11.6%, 0.1% et 90.1% repectivement dans le cas de la fusion des experts profil et image uniquement)(partie 2, section 4.5). Précisons les conditions rigoureuses dans lesquelles ces performances ont été obtenues et l'utilisation d'ensembles d'apprentissage et de test mutuellement exclusifs: un individu donné (tant client qu'imposteur) ne peut faire partie de ces deux ensembles simultanément. Cette condition est restrictive en ce qui concerne les données clients. Elle correspond à une situation où le superviseur aurait été optimisé en usine, faisant usage de clients différents de ceux qui utiliseront le système final. On peut s'attendre à des résultats meilleurs encore lors d'un apprentissage effectué sur site, impliquant les personnes mêmes qui utiliseront le système par la suite.

Perspectives futures

Les améliorations possibles du système présenté tout au long de ce travail, peuvent être de deux types: des améliorations d'ordre pratique nécessaires pour pouvoir passer d'un système en phase de développement à un système opérationnel, et des améliorations d'ordre plus conceptuel. Comme améliorations pratiques, nous pouvons citer entre-autres

- la prise en compte de fonds non uniformes et non statiques lors de la segmentation du visage;
- la sélection automatique d'une image de face ou de profil la plus représentative du client, pendant que celui-ci se présente devant le système;
- la prise en compte de modifications d'éclairage non uniformes, entre les images de référence et candidate;
- le développement d'une interface système/utilisateur conviviale, invitant par exemple l'utilisateur à se positionner correctement si tel n'est pas le cas.

En ce qui concerne les améliorations d'ordre conceptuel, nous pourrions envisager lors d'un travail futur

- la prise en compte de seuils individuels au sein du superviseur;
- la prise en compte d'une mesure de confiance ou de qualité de l'authentification au droit de chaque expert. Pour l'expert profil, cette mesure pourrait par exemple être liée à la longueur du profil sur lequel s'effectue l'authentification.

Décrivons enfin une application possible et plus que probable d'un système d'authentification biométrique tel que celui proposé ici. Elle consiste à crypter au sein d'une carte magnétique à code personnel, telle une carte de banque, les données biométriques du client. Celles-ci sont alors comparées avec les caractéristiques du détenteur de la carte lors de chaque utilisation. Le code secret resterait nécessaire afin de bénéficier d'un niveau de sécurité maximal face aux éventuels malfaiteurs[2]. En ayant pris soin de calibrer le système biométrique de telle façon à ce qu'il ne rejette aucun client, on peut estimer que notre superviseur linéaire parviendrait encore à rejeter de l'ordre de 99 imposteurs sur 100[3] dans le cadre de la fusion des experts profil, frontal et vocal. Ceci revient à faire baisser le taux d'imposture inhérent aux cartes magnétiques à code personnel d'un facteur 100. Ce résultat plus qu'encourageant laisse donc entrevoir l'avènement de telles techniques d'authentification dans un futur proche.

2. Un code à quatre chiffres correspond statistiquement à un taux de fausse acceptation de 0.01%. Ce taux est néanmoins plus élevé dans la pratique, les malfaiteurs cherchant en général à connaître le code secret des cartes qu'ils volent ensuite.

3. Extrapolation à partir du taux de succès mentionné au tableau 4.7.

Annexe A

Minimisation du Simplexe

L'algorithme du *simplexe*, aussi appelé algorithme de *Nelder et Mead* du nom de ses auteurs, est une méthode classique de minimisation d'une fonction à variable multidimensionnelle. Cet algorithme n'est pas le plus efficace en terme de coût de calcul, mais présente l'avantage de ne recourir qu'à des évaluations de la fonction à minimiser et non de ses dérivées. Pour des problèmes dont la complexité n'est pas trop grande, tel le cas de nos applications où les fonctions à minimiser sont définies sur des espaces de dimension 4 ou 5, l'algorithme du simplexe fournit des solutions précises en généralement moins d'une trentaine d'itérations. Cet algorithme se décrit aisément de façon graphique. Pour ce qui en est de son implémentation pratique, un code détaillé peut être obtenu en [54].

On appelle simplexe, un polyèdre convexe à $n+1$ sommets dans un espace à n dimensions, ici les n variables de la fonction $f(x_1, ..., x_n)$ que l'on désire minimiser. En deux dimensions, le simplexe est un triangle; en trois dimensions, un tétraèdre. Lors de l'initialisation de l'algorithme, l'utilisateur fournit des coordonnées pour chaque sommet, en prenant garde de donner au simplexe un volume suffisamment grand pour englober le minimum recherché. Après avoir généré le simplexe initial, l'algorithme procède à différentes déformations jusqu'à ce que l'ensemble des sommets se soit réduit en un seul point, le minimum recherché. La déformation la plus fréquente correspond à une *réflexion* et consiste à projeter le sommet du simplexe pour lequel $f(x_1, ..., x_n)$ est le plus élevé, à travers la base du côté opposé, comme illustré à la figure A.1 (a). Les autres déformations sont la *dilatation* et la *contraction* et sont représentées aux figures A.1 (b) et (c).

Il existe naturellement de nombreuses combinaisons possibles entre ces diverses déformations. En général, après une réflexion, et dans le cas où le nouveau sommet est meilleur que celui que l'on vient de quitter, l'algorithme tente une dilatation dans cette direction. Dès que l'évaluation de la fonction au droit du sommet dilaté devient plus faible que celle relative à l'un des sommets qui n'a pas encore été déplacé, on redémarre la procédure sur ce dernier. Si après un certain nombre d'itérations, on n'observe plus de gain appréciable, le simplexe est alors contracté dans son entièreté et la procédure du simplexe réinitialisée. Un telle séquence d'opérations aboutit toujours à cerner un minimum, mais celui-ci peut n'être qu'un minimum local.

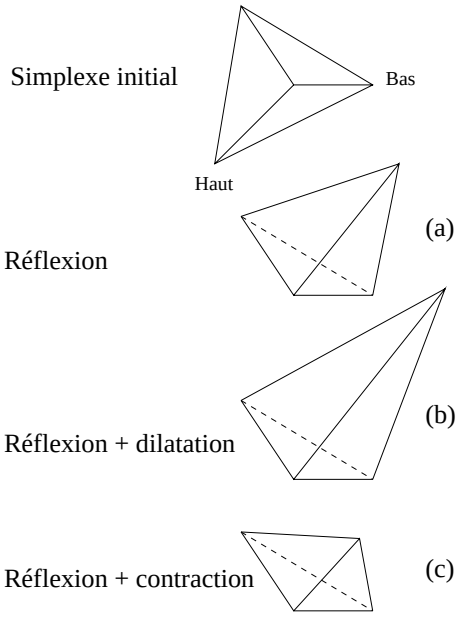


Figure A.1 - *Déformations possibles du simplexe*

Annexe B

TFA et TFR pour les opérateurs ET et OU

En reprenant les notations utilisées dans le cadre de la section 3.2.1, partie 2, soit:

- (fa_1, fr_1) et (fa_2, fr_2) deux points de fonctionnement pris sur les courbes caractéristiques de deux experts *indépendants*
- $fr_i = CC_i(fa_i)$ où CC_i désigne la fonction décroissante qui définit la courbe caractéristique de l'expert i $(i = 1, 2)$.
- $0 \leq fa_i, fr_i \leq 1$

on peut construire le tableau B.1. Celui-ci récapitule les probabilités d'acceptation et de rejet du candidat, après fusion, en fonction des décisions prises par les experts (acceptation/rejet), la classe du candidat (client/imposteur) et le type de superviseur considéré (ET/OU).

La probabilité de fausse acceptation dans le cas d'un superviseur basé sur un opérateur ET apparaît dans la première ligne de la quatrième colonne, tandis que pour l'opérateur OU, on sommera les lignes 1, 2, et 3 de cette même colonne (lignes pour lesquelles l'imposteur est effectivement accepté, voir colonne 6) :

$$FA_{et} \quad = \quad fa_1 fa_2$$

$$\begin{aligned} FA_{ou} &= fa_1 fa_2 + fa_1(1-fa_2) + (1-fa_1)fa_2 \\ &= fa_1 + fa_2 - fa_1 fa_2 \end{aligned}$$

La probabilité de faux rejet quant à elle, est obtenue par sommation des probabilités apparaissant aux lignes 2, 3 et 4 de la troisième colonne pour l'opérateur ET, et par la dernière ligne de cette colonne pour le OU.

$$\begin{aligned} FR_{et} &= (1-fr_1)fr_2 + fr_1(1-fr_2) + fr_1 fr_2 \\ &= fr_1 + fr_2 - fr_1 fr_2 \\ FR_{ou} &= fr_1 fr_2 \end{aligned}$$

Nous obtenons ainsi les équations de (3.3) à (3.6) introduites en début de section 3.2.

Décisions experts		Probabilité de l'événement selon la classe du candidat		Décisions superviseur	
Exp. 1	Exp. 2	Client	Imposteur	ET	OU
1	1	$(1-fr_1)(1-fr_2)$	$fa_1 fa_2$	1	1
1	0	$(1-fr_1)fr_2$	$fa_1(1-fa_2)$	0	1
0	1	$fr_1(1-fr_2)$	$(1-fa_1)fa_2$	0	1
0	0	$fr_1 fr_2$	$(1-fa_1)(1-fa_2)$	0	0

Tableau B.1 - *Récapitulatif de tous les cas d'acceptation et de rejet possibles selon les décisions des experts (1=acceptation, 0=rejet), la classe du candidat et le type de superviseur*

Bibliographie

[1] T. Aibara, K. Ohue, and Y. Matsuoka. "Human face recognition of p-type Fourier descriptors". In *Visual Communications and Image Processing (VCIP'91)*, pages 198–203. Proceedings of SPIE no 1606, 1991.

[2] A. Ayoun. "Fusion amont en détection de sources ponctuelles". In *Proceedings des Journées Thématiques en Fusion d'Information GDR-PRC ISIS*, pages 69–74, Paris, France, October 1997.

[3] C. Beumier and M. Acheroy. "Automatic profile identification". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 145–152, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[4] E. S. Bigün, J. Bigün, B. Duc, and S. Fischer. "Expert conciliation for multi modal person authentication systems by bayesian statistics". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 291–300, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[5] W. W. Bledsoe. "Man-machine facial recognition". Panoramic Research Inc. Report no PRI22, Palo Alto, CA, 1966.

[6] P. Bolleire. "Reconnaissance de visages par vue frontale". Mémoire de fin d'études, Université catholique de Louvain, June 1997.

[7] G. Borgefors. "Distance transformations in digital images". *Computer Vision, Graphics, Image Processing*, Vol. 34, no. 3, pp. 344–371, June 1986.

[8] J. D. Broffitt. "Nonparametric classification". In P. R. Krishnaiah and L. N. Kanal, editors, *Handbook of statistics 2. Classification, pattern recognition and reduction of dimentionality.* North-Holland, 1982.

[9] R. Brunelli and D. Falavigna. "Person identification using multiple cues". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, no. 9, pp. 955–966, September 1995.

[10] R. Brunelli and T. Poggio. "Face recognition through geometrical features". In *Proceedings of the Second European Conference on Computer Vision (ECCV'92)*, pages 792–800, S. Margherita, Ligure, Italy, May 1992. Springer Lecture Notes in Computer Science (LNCS) no 588.

[11] R. Brunelli and T. Poggio. "Face recognition: features versus templates". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, no. 10, pp. 1042–1052, October 1993.

[12] M. A. Butt and P. Maragos. "Optimum design of chamfer distance transforms". *IEEE Transactions on Image Processing*, Vol. 7, no. 10, pp. 1477–1484, 1997.

[13] R. Chellappa, C. L. Wilson, and S. Sirohey. "Human and machine recognition of faces: a survey". *Proceedings of the IEEE*, Vol. 83, no. 5, pp. 705–740, May 1995.

[14] G. Chollet and F. Bimbot. "Assessment of speaker verification systems". In *Spoken Language Resources and Assessment*. EAGLES Handbook, 1995.

[15] G. Chollet, J.-L. Cochard, A. Constantinescu, and P. Langlais. "Swiss French polyphone and polyvar: telephone speech databases to study intra and inter speaker varaibility". Technical report, IDIAP, Martigny, Switzerland, 1995.

[16] I. Craw, H. Ellis, and J. R. Lishman. "Automatic extraction of face features". *Pattern Recognition Letters*, Vol. 5, no. 2, pp. 183–187, February 1987.

[17] J. R. Deller, J. G. Proakis, and J. H. L. Hansen. *Discrete-time processing of speech sugnals*. Macmillan Publishing Company, 1993.

[18] P. Delogne. "A propos des récepteurs optimaux". *Revue H.F.*, Vol. 8, no. 3, pp. 61–65, 1970.

[19] B. Duc, E. S. Bigün, J. Bigün, G. Maître, and S. Fischer. "Fusion of audio and video information for multi modal person authentication". *Pattern Recognition Letters*, Vol. 18, no. 9, pp. 835–843, September 1997.

[20] B. Duc, S. Fisher, and J. Bigün. "Face authentication with sparse grid gabor information". In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, pages 3129–3132, Munich, Germany, April 1997.

[21] B. Duc, G. Maître, S. Fisher, and J. Bigün. "Person authentication by fusing face and speech information". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 311–318, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[22] B. Efron and R. Tibshirani. "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy". *Statistical Science*, Vol. 1, no. 1, pp. 54–77, 1986.

[23] S. Fischer and B. Duc. "Shape normalization for face recognition". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 21–26, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[24] R. A. Fisher. "The use of multiple measurements in taxonomic problems". *Annals of Eugenics*, Vol. 7, pp. 179–188, 1936.

[25] S. Furui. "Cepstral analysis technique for automatic speaker verification". *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. 29, no. 2, pp. 254–272, 1981.

[26] S. Furui. "Recent advances in speaker recognition". *Pattern Recognition Letters*, Vol. 18, no. 9, pp. 859–872, September 1997.

[27] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Osu. "Face recognition system using local autocorrelations and multiscale integration". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, no. 10, pp. 1024–1028, October 1996.

[28] L. D. Harmon and W. F. Hunt. "Automatic recognition of human face profiles". *Computer Graphics and Image Processing*, Vol. 6, pp. 135–156, 1977.

[29] L. D. Harmon, M. K. Khan, R. Larsh, and P. F. Raming. "Machine identification of human faces". *Pattern Recognition*, Vol. 13, no. 2, pp. 97–110, 1981.

[30] S. Haykin. *Digital communications*. Wiley, 1988.

[31] F. Janez. "Généralités sur la fusion de données en reconnaissance". In *Proceedings des Journées Thématiques en Fusion d'Information GDR-PRC ISIS*, pages 159–163, Paris, France, October 1997.

[32] X. Jia and M. S. Nixon. "Extending the feature set for automatic face recognition". In *IEE 4th Conference on Image Processing and Applications*, pages 155–158, Maastricht, The Netherlands, April 1992.

[33] P. Jourlin, J. Luettin, D. Genoud, and H. Wassner. "Acoustic-labial speaker verification". *Pattern Recognition Letters*, Vol. 18, no. 9, pp. 853–858, September 1997.

[34] T. Kanade. *Computer recognition of human faces*. Volume 47 of Interdisciplinary System Research, Birkhauser Verlag, Basel and Stuttgart, 1977.

[35] M. Kirby and L. Sirovich. "Application of the Karhunen-Loève procedure for the characterization of human faces". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, no. 1, pp. 103–108, January 1990.

[36] J. Kittler, J. Matas, K. Jonsson, and R. Sanchez. "Combining evidence in personal identity verification systems". *Pattern Recognition Letters*, Vol. 18, no. 9, pp. 845–852, September 1997.

[37] W. Konen and E. Schulze-Krüger. "Zn-face: a system for access control using automated face recognition". In *International Workshop on Automatic- and Gesture Recognition*, pages 18–23, Zurich, Switzerland, June 1995.

[38] C. Kotropoulos, I. Pitas, S. Fischer, and B. Duc. "Face authentication using morphological dynamic link architecture". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 169–176, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[39] P. Kruizinga and N. Petkov. "Optical flow applied to person identification". In *Proceedings of the 1994 EUROSIM Conference on Massively Parallel Processing Applications and Development*, pages 871–878, Delft, The Netherlands, June 1994. Elsevier, Amsterdam.

[40] M. Kunt, G. Granlung, and M. Kocher. *Traitement de l'information (Volume 2): Traitement numérique des images.* Presses Polytechniques et Universitaires Romandes, 1993.

[41] K.-M. Lam and H. Yan. "An analytic-to-holistic approach for face recognition based on a single frontal view". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, no. 7, pp. 673–686, July 1998.

[42] J. Makhoul and R. Schwartz. "The voice of the computer". IEEE Spectrum, December 1997.

[43] J. Matas and J. Kittler. "Spatial and feature space clustering: applications in image analysis". In *6th International Conference on Computer Analysis and Patterns*, Prague, Czech Republic, September 1995.

[44] B. Moghaddam and A. Pentland. "Face recognition using view-based and modular eigenspaces". In *Automatic Systems for the Identification and Inspection of Humans, SPIE Vol. 2277*, July 1994.

[45] A. Pentland, T. Starner, N. Etcoff, A. Masoiu, O. Oliyide, and M. Turk. "Experiments with eigenfaces". In *Looking at People Workshop, The International Joint Conference on Artificial Intelligence (IJCAI'93)*, Chamberry, France, August 1993.

[46] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi. "The FERET Septembre 1996 database abd evaluation procedure". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 395–402, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[47] P. J. Phillips and Y. Vardi. "Data-driven methods in face recognition". In *International Workshop on Automatic- and Gesture Recognition*, pages 65–70, Zurich, Switzerland, June 1995.

[48] S. Pigeon. "The M2VTS multimodal face database (release 1.00)". European ACTS Deliverable AC102/UCL/WP1/DS/P/161, December 1996.

[49] S. Pigeon and L. Vandendorpe. "The M2VTS multimodal face database (release 1.00)". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 403–409, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[50] S. Pigeon and L. Vandendorpe. "Profile authentication using a chamfer matching algorithm". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, pages 185–192, Crans-Montana, Switzerland, March 1997. Springer Lecture Notes in Computer Science (LNCS) no 1206.

[51] S. Pigeon and L. Vandendorpe. "Profile authentication using a chamfer matching algorithm". In *Proceedings of the Eleventh Conference on Quantitative Methods for Decision Making (ORBEL11)*, Namur, Belgium, January 1997.

[52] S. Pigeon and L. Vandendorpe. "Image-based multimodal face authentication". *Signal Processing*, Vol. 69, no. 1, pp. 59–79, August 1998.

[53] S. Pigeon and L. Vandendorpe. "Multiple experts for robust face authentication". In *Optical Security and Counterfeit Deterrence Techniques II*, pages 166–177, San Jose, California, January 1998. Proceedings of SPIE no 3314.

[54] W. H. Press, S. A. Teukolsky, and W. T. Vertterling. *Numerical recipes in C: The art of scientific computing.* Cambridge University Press, 1988.

[55] B. Ramaekers. "Détermination de la position angulaire instantanée d'un visage dans une séquence en mouvement". Mémoire de fin d'études, Université catholique de Louvain, June 1997.

[56] A. Rosenfeld and J.L. Pfaltz. "Sequential operations in digital picture processing". *Journal of the ACM*, Vol. 13, no. 4, pp. 471–494, October 1966.

[57] J. C. Russ. *Image processing handbook.* IEEE Press, 1995.

[58] T. Sakai, M. Nagao, and M. Kidode. "Processing of multilevel pictures by computer – the case of photographs of human faces". *System, Computers, Controls*, Vol. 2, no. 3, pp. 47–53, 1977.

[59] F. S. Samaria and A. C. Harter. "Parametrization of a stochastic model for human face identification". In *2nd IEEE Workshop on Applications of Computer Vision*, Sarasosa, Florida, December 1994.

[60] J.-C. Simon. "La reconnaissance des formes à l'épreuve des faits". *La recherche*, Vol. 312, pp. 58–62, September 1998.

[61] J.-P. Thiran. *Représentation et recalage d'images tridimensionnelles par squelettes morphologiques.* PhD thesis, Université catholique de Louvain, July 1997.

[62] H. L. Van Trees. *Detection, Estimation and Modulation Theory.* Wiley, 1968.

[63] S. Valente and J.-L. Dugelay. "A multi-site teleconferencing system using V.R. paradigms". In *Proceedings of the 2nd European Conference on Multimedia Applications, Services and Techniques (ECMAST'97)*, pages 359–374, Milan, Italy, May 1997. Springer Lecture Notes in Computer Science no 1242.

[64] P. Verlinde and G. Chollet. "Combining vocal and visual cues in an identity verification system using k-nn based classifiers". In *Proceedings of the IEEE Workshop on Multimedia Signal Processing*, Los Angeles, CA, December 1998.

[65] P. Verlinde, D. Genoud, G. Gravier, and G. Chollet. "Proposition d'une stratégie de fusion de données à trois niveaux pour la vérification d'identité". XXXIIIèmes journées d'études de la parole, Martigny, Switzerland, June 1998.

[66] R. Viswanathan and P. K. Varshney. "Distributed detection with multiple sensors". *Proceedings of the IEEE*, Vol. 85, no. 1, pp. 54–63, January 1997.

[67] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. "Face recognition and gender determination". In *International Workshop on Automatic- and Gesture Recognition*, pages 92–97, Zurich, Switzerland, June 1995.

[68] K. Yu, X. Jiang, and H. Bunke. "Face recognition by facial profile analysis". In *International Workshop on Automatic- and Gesture Recognition*, pages 208–213, Zurich, Switzerland, June 1995.

[69] A. L. Yuille. "Deformable templates for face recognition". *Journal of Cognitive Neuroscience*, Vol. 3, no. 1, pp. 59–70, January 1991.