Vérification de l'Identité par Les Données Biométriques

Van-Bao Ly, R. Blouet, S. Renouard, S. Garcia-Salicetti, B. Dorizzi, G. Chollet

Abstract: **This article presents the main idea of a Biometric Authentication System (BAS). We will briefly describe BIOMET database, constructed for this purpose and composed of five modalities: face, hand-shape, fingerprint, online signature and speech. We have used two user-friendly biometric modalities of this database to construct the unimodal verification systems based on online signature and speech. The Signature Verification system relies on Hidden Markov Models (HMMs), and two kinds of Speaker Verification systems have been designed. The first one is text-dependent, using Dynamic Time Warping (DTW) to compute a decision score. The second one is text-independent and use Gaussian Mixture Models (GMMs) to compute the likelihood ratio of an utterance. We present also two classical learning-based fusion techniques: an additive CART-trees classifier built with boosting and Support Vector Machines (SVMs). In particular, the signature modality was fused with clean and noisy speech, at two different levels of degradation. We have at our disposal 68 real persons, divided into two equal bases BAF and BTF to train and test the unimodal systems as well as the fusion systems. The performances of the systems are evaluated on 20 different couples (BAF, BTF), randomly chosen from 68 original real persons. In the non noisy case, we have dropped down the total error rate to 2.21% from 7.65% for the best unimodal system**.

Keywords: **Additive Tree Classifier, Dynamic Time Warping, Fusion, Hidden Markov Models, Gaussian Mixture Models, Multimodal Identity Verification, Support Vector Machines.**

I. INTRODUCTION

La vérification de l'identité est très importante dans la vie quotidienne. En fait, notre identité est vérifiée lorsque nous entrons dans notre lieu de travail, lorsque nous nous connectons au réseau informatique, lorsque nous exécutons des transactions bancaires, etc…

Il y a deux manières classiques de vérifier l'identité d'un individu. L'une est basée sur une *connaissance*, par exemple un mot de passe, et l'autre est basée sur une *possession*, par exemple une pièce d'identité, une clé, un badge... Par fois, ces deux manières sont utilisées en parallèle; c'est par exemple le cas d'une carte à puce avec un code confidentiel.

Pourtant, les *possessions* peuvent être volées ou perdues et les *connaissances* peuvent être oubliées… Ainsi, la biométrie représente une alternative à ces faiblesses, pour vérifier l'identité d'une personne. En effet, elle consiste à utiliser des caractéristiques *physiques* d'un individu comme son visage ou ses empreintes digitales, ou bien certaines caractéristiques *comportementales* comme sa signature manuscrite. La biométrie est ainsi reliée à la personne, très difficile à mimer, et elle ne peut jamais être perdue ou volée.

Cependant, les caractéristiques biométriques dépendent beaucoup de l'environnement de capture, du stress de l'individu ou de son état général, ce qui gène le bon fonctionnement du système de vérification biométrique. Ainsi, nous pouvons fiabiliser la performance du système biométrique en utilisant simultanément plusieurs modalités différentes [1, 2, 3, 4, 5, 6, 7, 8]. De plus, l'utilisation de plusieurs modalités permet de bâtir un système plus souple vis à vis des situations rencontrées: par exemple, lorsqu'une modalité est défaillante, le système peut reposer sur d'autres modalités en accordant peu de confiance à la modalité défaillante.

Dans cet article, nous présenterons d'abord la base BIOMET [9], qui contient cinq modalités: la signature en ligne, la parole, le visage, l'empreinte digitale et la forme de la main. Cette base a été construite dans l'objectif d'étudier l'apport de la fusion multimodale. Dans ce travail, deux modalités de la base ont été utilisées pour cette étude, la signature en ligne et la parole qui seront ensuite brièvement décrites. Puis nous présenterons le système de vérification de signatures, basé sur les Modèles de Markov Cachés (Hidden Markov Models - HMMs) [10], et les deux systèmes de vérification de parole, l'un dépendant du texte, fondé sur une distance élastique (Dynamic Time Warping - DTW) [11] et l'autre indépendant du texte, utilisant le Modèle de Mélange de Gaussiennes (Gaussian Mixture Models - GMMs) [12]. Ensuite, deux méthodes de fusion sont décrites: une méthode de Classification Arborescente Additive (Additive Tree Classifier -ATC) qui exploite un critère de théorie de l'information [13] dans le cadre du "Boosting" [5, 14, 15], et une méthode d'apprentissage statistique utilisé avec succès en vérification de l'identité, les Machines à Vecteurs de Support (Support Vector Machines - SVM) [1, 2, 3, 7]. Finalement, les résultats des systèmes unimodaux et des systèmes de fusion sont analysés.

II. BASE DE DONNÉES BIOMET

Depuis quelques années, les Ecoles du GET (Groupe des Ecoles de Télécommunication) ont développé de façon indépendante diverses méthodes d'identification et d'authentification en traitement de la parole (identification

B. Ly-Van, S. Garcia-Salicetti, B. Dorizzi, Institut National des Télécommunication, Dépt. EPH, 9 rue Charles Fourier, 91011 EVRY France. Email: {Bao.Ly_van, Sonia.Salicetti, Bernadette.dorizzi}@int-evry.fr.

R. Blouet, S. Renouard, G. Chollet, Ecole Nationale Supérieure des Télécommunications, Lab. CNRS-LTCI, 46 rue Barrault, 75634 Paris. Email: {Blouet, Renouard, Chollet}@tsi.enst.fr

du locuteur - ENST Paris), de l'image (analyse d'empreintes digitales, de visage - équipe EURECOM) et de l'écriture (authentification de signatures - INT). Il est de plus apparu au sein de ces mêmes Ecoles un savoir-faire tout à fait conséquent dans le domaine de la fusion d'informations. Ce contexte a ainsi naturellement donné lieu à une action concertée de ces divers acteurs en vue du thème porteur de la vérification biométrique multimodale d'identité.

En effet, l'objectif du projet BIOMET est de mettre en synergie les compétences des équipes des Ecoles du GET impliquées dans le domaine de la vérification et de l'authentification pour accéder à un système sécurisé au moyen de diverses modalités biométriques: vérification de signatures, reconnaissance du visage, des empreintes digitales et de la forme de la main, authentification du locuteur, ainsi que leur fusion pour la vérification multimodale de l'identité. Les tâches principales du projet sont:

- Créer une base de données multimodale contenant une centaine de personnes environ, impérative préalable à tout travail en commun sur la fusion des différentes modalités mentionnées.
- Mettre au point des systèmes de vérification opérationnels pour chacune des modalités.
- Chercher une ou plusieurs stratégies de fusion adéquates, et les tester sur cette plate-forme.

Concrètement, la base BIOMET, acquise en 2002 sur 3 campagnes espacées de quelques mois, contient 5 modalités biométriques: la signature en ligne, la parole, la forme de la main, l'empreinte digitale et le visage, et cela pour environ 100 personnes. Nous présentons à la suite en détail le contenu de la base pour les deux modalités concerné dans cet article: la signature en-ligne et la parole.

A. Signature en-ligne

Les signatures de la base BIOMET ont été acquises sur une tablette à digitaliser WACOM Intuos2 A6, avec un stylo à encré (inkpen), à la fréquence d'échantillonnage de 200Hz. Une signature est représentée par une suite de points. A chaque point, la tablette capture les 5 paramètres suivants: les coordonnées (*x, y)*, la pression axiale *p* qui dépend beaucoup de signataire et les deux angles *azimut* et *altitude*, représentant la position spatiale du stylo.

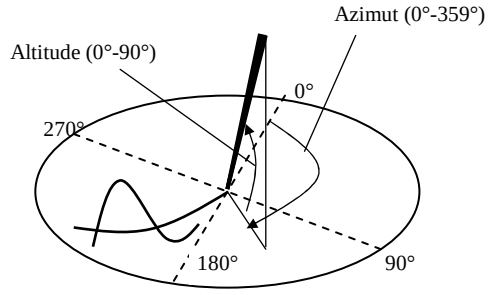

Fig. 1. Les angles azimut et altitude

Les signatures d'une personne sont capturées en deux sessions de 5 mois de délai. Lors de la première session, 5 signatures vraies et 6 signatures imitées ont été capturées par personne. Lors de la deuxième session, 10 signatures vraies et 6 signatures imitées ont été saisies par personne. Les 12 signatures imitées résultantes d'une personne ont été réalisées par 4 imitateurs différents, chacun en a fait 3. Les imposteurs essaient de mimer l'image de la signature vraie.

Après une dernière sélection, la base contient finalement les signatures de 84 personnes.

B. Parole

Les données de parole sont acquises en deux sessions différentes en utilisant le même microphone d'une caméra numérique. Un délai de 3 mois est assuré entre deux sessions. Le signal est quantifié à 16KHz, et chaque échantillon est codé par 16 bits. Pour les deux sessions, chaque personne prononce les 10 chiffres, dans l'ordre ascendant et descendant. Les données disponibles d'une personne sont d'environ 90 secondes par session.

Comme nous voulons tester la fusion sur une base bimodale (signature et parole) de vraies personnes, nous n'avons que 68 personnes qui possèdent à la fois les données de signature et de la parole. Dans la partie 5, nous réalisons des expériences pour tester les systèmes unimodaux ainsi que les systèmes de fusion sur ces 68 personnes de la base.

III. SYSTÈMES BIOMÉTRIQUES DE VÉRIFICATION UNIMODAUX

La figure 2 présente un système de vérification biométrique typique. Le module d'apprentissage capture les

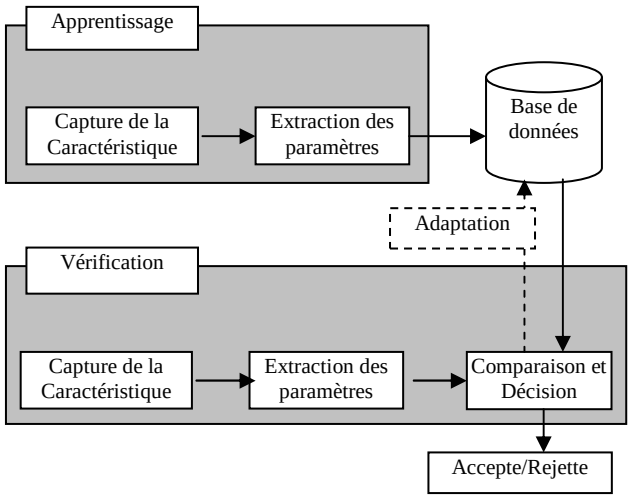

Fig. 2. Système de vérification biométrique général

caractéristiques de la personne. Pourtant, seulement les paramètres les plus représentatifs sont extraits et stockés dans la base de données. Le modèle de l'utilisateur, comme le HMM pour la signature, ou le GMM pour la parole, est une représentation compacte des données. Au lieu de stocker les paramètres biométriques d'une personne, le système sauvegarde les paramètres de son modèle. Dans le cas d'un GMM, ces paramètres sont les matrices de covariance et les moyennes de chaque gaussienne composant le mélange. Cette technique permet de diminuer les données stockées ainsi que faciliter la phase de vérification. Parfois, les données de la phase de vérification déjà authentifiées sont utilisées pour ajuster les paramètres du modèle de l'utilisateur.

Rappelons qu'un modèle est seulement un *estimateur* qui donne, pour un accès quelconque, un score. Pour construire un *classifieur*, il faut avoir un *seuil* en plus. Normalement, nous utilisons une base propre *BA* contenant un nombre suffisant d'exemples vrais et imités de quelques dizaines d'utilisateurs pour calculer ce seuil, et nous espérons que ce seuil soit robuste pour le reste des utilisateurs potentiels. Il y a deux critères pour calculer le seuil: l'erreur totale (*ET*) minimum, et le taux d'erreur égal (*TEE*). *ET* minimum vise à chercher un seuil qui minimise l'erreur totale (*ET*) du système où l'erreur totale est définie comme étant:

$$ET=(\#FA + \#FR)/(\#V + \#I)$$

Avec *#FA* le nombre de fausse acceptation, *#FR* le nombre de faux rejet, *#V* le nombre d'accès client et *#I* le nombre d'accès imposteur. *TEE* vise à chercher un seuil qui

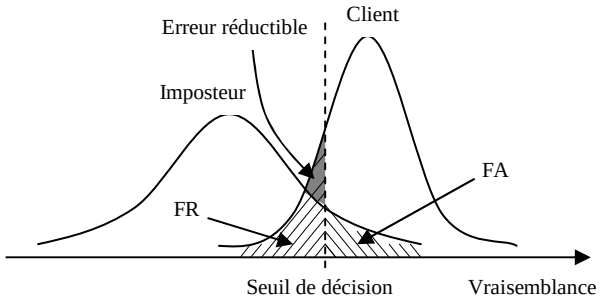

Fig. 3. Seuil de décision et les taux d'erreur

équilibre le taux de fausse acceptation *(FA)* et le taux de faux rejet *(FR)* sur la base *BA*, où *FA=#FA/#I* et *FR=#FR/#V*. Les taux d'erreur *FA*, *FR* et le seuil sont illustrés dans la figure 3.

Nous présentons à la suite les systèmes de vérification unimodaux que nous avons construits au sein du projet BIOMET.

A. *Vérification de signature en-ligne*

1) *Pré-traitement et encodage de la signature*:

Les données d'une signature contiennent du bruit, généré par la haute fréquence d'échantillonnage et, d'autre part, par la quantification par la tablette des 5 paramètres acquis. Nous avons utilisé différentes stratégies pour filtrer les paramètres, comme décrite dans [7]. A partir des coordonnées, nous pouvons extraire d'autres paramètres comme la vitesse, l'accélération… Finalement, nous avons 12 paramètres dynamiques à chaque point d'une signature.

2) *Modélisation des signatures:*

Nous modélisons les signatures d'une personne par un modèle HMM continu [10, 16]. Comme nous disposons de peu de signatures par personne pour entraîner son modèle, nous utilisons la méthode "bagging" [15, 17] pour construire le modèle HMM "agrégé" (en anglais "aggreated" HMM). Cela consiste en fait à générer, à partir des N signatures disponibles, T ensembles de N signatures en tirant au hasard avec remise. Avec ces T ensembles, nous entraînons T modèles HMM. Le modèle HMMs "aggrégé" obtenu exploite les scores de ces T modèles HMM de façon suivante: pour une signature O quelconque, la moyenne des T scores de O sur les T modèles HMMs est considérée comme le score "composite" S(O) du modèle HMM "aggrégé" sachant O.

Pour authentifier une signature *O* de la personne *i*, nous calculons la différence absolue entre le score "composite" $S_i(O)$ et la moyenne S_i^* des *N*T* scores de toutes les signatures des *T* ensembles, donnés par les *T* modèles associés (S_i^* est appelé la moyenne des scores de la base d'apprentissage), puis nous comparons cette différence avec un seuil. La signature *O* du signataire *i* est authentifiée si et seulement si:

$$| S_i(O) - S_i^* | < \tau \qquad (1)$$

où τ est le seuil global, calculé pour tous les signataires en utilisant le principe précédemment mentionné.

B. *Vérification de parole*

Dans un système de vérification de locuteur, nous avons deux hypothèses pour une donnée parole *X*:

H_λ : *X* est prononcée par λ et

$H_{\bar{\lambda}}$: *X* est prononcée par une autre personne que λ.

La décision d'acceptation ou de rejet de l'identité proclamée est faite de la façons suivante:

$$\log \frac{D_\lambda(X)}{D_{\bar{\lambda}}(X)} \begin{cases} \geq \beta & \text{on accepte } \lambda \\ < \beta & \text{on rejette } \lambda \end{cases} \qquad (2)$$

où $D_\lambda(X)$ et $D_{\bar{\lambda}}(X)$ sont respectivement les mesures de similarité de *X* sachant H_λ et $H_{\bar{\lambda}}$, et β est le seuil de décision, calculé avec le même principe énoncé pour le calcul du seuil du système de signature.

Dans notre système de vérification du locuteur *dépendant du texte*, la Distance Elastique (Dynamic Time Warping - DTW) [11] est utilisée pour calculer $D_\lambda(X)$ et $D_{\bar{\lambda}}(X)$. Pour le système de vérification du locuteur *indépendant du texte*, $D_\lambda(X)$ et $D_{\bar{\lambda}}(X)$ correspondent aux valeurs de la fonction de densité de probabilité P_λ et $P_{\bar{\lambda}}$ sachant *X*. Nous avons utilisé un Modèle de Mélange de Gaussiennes (Gaussian Mixture Model - GMM) [12] pour estimer ces fonctions de densité.

1) *Vérification de locuteur dépendant du texte:*

La mesure de similarité $D_\lambda(X)$ dans l'équation (2) est estimée par la distance *DTW* [11] entre les données d'apprentissage X_λ de 4 chiffres et la donnée de test *X* des mêmes 4 chiffres (plus la distance est grande, plus la mesure de similarité est petite). Comme décrite dans [18], nous utilisons un groupe de locuteurs pour calculer $D_{\bar{\lambda}}(X)$. Le modèle du monde $\overline{\lambda}$ est représenté par *K* exemplaires $\{X_1..X_K\}$ des mêmes chiffres prononcés par *K* locuteurs différents. La mesure $D_{\bar{\lambda}}(X)$ est estimée par la moyenne des distances *DTW* entre *X* et X_k où $k = 1..K$. Pour les expérimentations de cet article, nous avons choisi *K=50* pour le modèle du monde.

2) *Vérification de locuteur indépendant du texte:*

Dans le système de vérification de locuteur indépendant du texte, nous utilisons un seul modèle du monde $P_{\bar{\lambda}}(X)$. Ce modèle, aussi appelé le Modèle Universel (Universal

Background Model - UBM) [19], est un GMM à 256 gaussiennes et ses matrices de covariance sont diagonales. Le modèle du client $P_\lambda(X)$ est obtenu en adaptant le Modèle Universel sur les données d'apprentissage correspondantes. Le modèle du monde $P_{\bar{\lambda}}(X)$ et le modèle du client $P_\lambda(X)$ sont utilisés pour calculer le score d'une donnée à la place de $D_{\bar{\lambda}}(X)$ et $D_\lambda(X)$ respectivement dans l'équation (2).

IV. Systèmes de fusion

Nous pouvons désormais construire un système de fusion utilisant les 3 scores des 3 experts unimodaux, de façon normalisée comme suit: la première entrée, issue du système de signature:

$$(S_i(O) - S_i^*)/\sigma$$

où $S_i(O)$ et S_i^* sont décrits dans l'équation (1) et σ est la moyenne des σ_i, σ_i étant l'écart-type des scores des $N*T$ signatures des T ensembles d'apprentissage, comme décrit dans 3.1.2, chacun donné par le modèle HMM correspondant de la personne i.

La deuxième et la troisième entrées sont les mesures $\log \frac{D_\lambda(X)}{D_{\bar{\lambda}}(X)}$ définies dans l'équation (2), avec les 2 types de vérification du locuteur: dépendant du texte et indépendant du texte.

Nous présentons à la suite deux systèmes de fusion que nous avons construits.

A. Classification Arborescente Additive

Le Boosting [14] permet de construire un Modèle Additif efficace à partir d'un modèle donné, dans ce cas un Arbre de Décision (Classification and Regression Tree - CART) [13] binaire. Le principe de CART est de diviser récursivement l'espace des observations, dans notre cas l'espace *3-D* des scores des systèmes unimodaux, c'est à dire des vecteurs $s=[s_1, s_2, s_3]$, en deux sous-espaces.

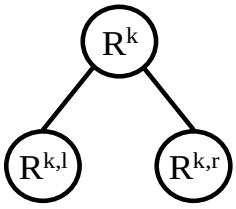

Fig. 4. Arbre de Décision

En effet, l'espace des observations R^k est divisé en deux sous-espaces $R^{k,l}$ et $R^{k,r}$ en maximisant ΔH:

$$\Delta H = H(R^k) - p_l H(R^{k,l}) - p_r H(R^{k,r})$$

où $H(R^k)$, $H(R^{k,l})$, $H(R^{k,r})$ sont respectivement les entropies des feuilles R^k, $R^{k,l}$ et $R^{k,r}$; $p_l = N^{k,l}/N^k$, $p_r = N^{k,r}/N^k$ où N^k, $N^{k,l}$ et $N^{k,r}$ sont respectivement le nombre d'observations dans les feuilles R^k, $R^{k,l}$ et $R^{k,r}$. L'entropie d'une feuille R est calculée par:

$$H(R) = p_\lambda(R).\log(p_\lambda(R)) + p_{\bar{\lambda}}(R).\log(p_{\bar{\lambda}}(R))$$

où $p_\lambda(R) = N_\lambda(R)/N(R)$ et $p_{\bar{\lambda}}(R) = 1 - p_\lambda(R)$

avec $N_\lambda(R)$ le nombre d'observations de la classe λ dans R et $N(R)$ le nombre d'observations total dans R.

Dans notre système, une feuille R^k est seulement divisée quand $N^k > 50$. Pour chaque vecteur $s=[s_1, s_2, s_3]$, nous allons chercher la région R de s classée par CART. Le score du vecteur s est alors $\log \frac{p(\lambda|s)}{p(\bar{\lambda}|s)}$, où $p(\lambda|s) = p_\lambda(R)$ et $p(\bar{\lambda}|s) = p_{\bar{\lambda}}(R)$.

Nous avons utilisé l'algorithme RealAdaboost [20] pour ajuster le Modèle Additif. Dans cet algorithme itératif, nous entraînons tout d'abord un Arbre de Décision CART. Puis les données d'apprentissage qui ne sont pas correctement classées par l'Arbre de Décision CART précédemment entraîné, sont retirées et pondérées d'une probabilité plus élevée que celles qui sont correctement classées afin de construire une nouvelle base d'apprentissage. Nous répétons cette itération pour construire plusieurs Arbres de Décision CART. Le score d'un vecteur s, donné par le système de fusion Classification Arborescente Additive est la moyenne des scores donnés par tous les Arbres de Décision CART.

B. Machines à vecteur de support

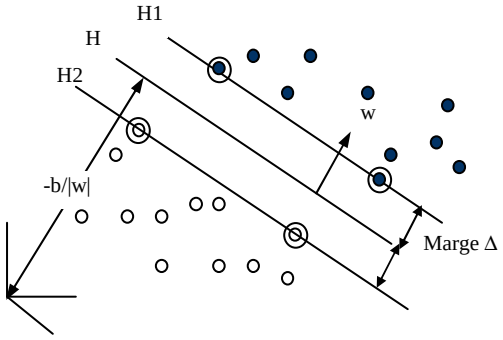

Fig. 5. Hyperplan H, données et vecteurs de support (encerclés) dans l'espace F.

Le principe des SVMs repose sur la recherche d'une séparatrice linéaire dans un espace de grande dimension qu'on construit parce que les données d'entrée ne sont pas linéairement séparables dans l'espace d'origine. Nous maximisons alors la marge, grandeur mesurant l'écart du modèle aux données, ce qui assure alors en principe de bonnes performances en généralisation. Soit $X = \{x_i\}$ l'ensemble des données étiquetées suivant $Y = \{y_i\}$ où $y_i = +1$ ou -1 qui représente la classe de chaque individu et Φ la fonction qui envoie les données d'entrée dans l'espace F de grande dimenssion, dit espace des caractéristiques. Comme défini sur la figure 5, la distance entre l'hyperplan:

$$H(w,b) = \{x \in F : \langle w , x \rangle + b = 0\}$$

et X est appelée la marge Δ. Avec le principe de Minimisation du Risque Structurel (SRM), Vapnik [21] a montré que maximiser la marge (ou minimiser $\|w\|$) permet de minimiser la VC-dimension du modèle de séparation, qui est un critère de généralisation efficace. Nous définissons alors le noyau K dans F:

$$K(x,y) = \langle \Phi(x), \Phi(y) \rangle$$

Cette fonction est en fait un raccourci mathématique qui nous évite de manipuler directement les éléments de F. Nous trouvons alors l'hyperplan optimal en résolvant, comme démontré dans [21], un problème convexe

quadratique. A partir des conditions d'optimalité de Karush-Kuhn-Tucker [21], on peut réécrire *w* suivant la décomposition parcimonieuse :

$$w = \Sigma_{i \in SV} \alpha_i y_i \Phi(x_i)$$

où $SV = \{i: \alpha_i > 0\}$ est l'ensemble des vecteurs de support.

Le choix de Φ, ou de manière équivalente *K*, est très important pour l'efficacité de la solution trouvée. On choisit ainsi traditionnellement le noyau polynômial de Vapnik $K(x,y)=<\Phi(x),\Phi(y)>^d$ ou le noyau Gaussien $K(x,y)=exp(-\gamma||x-y||^2)$. Nous avons choisi dans un premier temps un noyau linéaire (d = 1). En effet, l'utilisation de ce type de noyaux dans un cas similaire [1] de fusion, a donné de meilleures performances comparativement à d'autres choix.

Nous allons fusionner les scores des 3 experts unimodaux, chacun conçu pour une même personne. Ainsi, la dimension de l'espace d'origine est de 3.

V. Expérimentation

A. Protocole d'expérimentation

Nous disposons des données de 68 personnes réelles, chacune comptant 5 accès client bimodaux et 12 accès imposteur bimodaux qui n'ont jamais servi à l'entraînement des modèles de locuteur ou de signataire. Ces 68 personnes sont divisées en deux bases de 34 personnes *BAF* (Base d'Apprentissage de Fusion) et *BTF* (Base de Test de Fusion).

Le seuil τ du système de signature est estimé, comme décrit en 3.1.2, sur la partie signature de *BAF* et le seuil β du système de parole (dépendant et indépendant du texte) est estimé, comme décrit dans 3.2, sur la partie parole de *BAF*. Le critère *ET* minimum décrit dans la partie 3 est utilisé. Les taux d'erreur des systèmes unimodaux sont calculés sur les modalités correspondantes de *BTF*.

Les données bimodales dans *BAF* sont aussi utilisées pour entraîner les systèmes de fusion dont les taux d'erreur sont estimés sur la base *BTF*.

Avec ce protocole de test, nous pouvons voir facilement l'amélioration des performances des systèmes de fusion par rapport à celle des systèmes unimodaux. Pourtant, ce test est biaisé à cause du petit nombre de personnes ainsi que du petit nombre d'accès par personne. Pour surmonter ce problème, nous construisons plusieurs couples (*BAF, BTF)* différents, en tirant sans remise 34 personnes de *BAF*, le restant dédié à *BTF*. Les taux d'erreur "réels" des systèmes (unimodaux ou de fusion) sont estimés par la moyenne des taux d'erreur calculés sur plusieurs bases *BTF*. Les intervalles de confiance [22] des taux d'erreur moyens sont aussi calculés.

Comme les systèmes de vérification du locuteur sont beaucoup gênés par le bruit, nous avons dégradé les données parole pour voir la robustesse de notre deux systèmes de vérification du locuteur contre le bruit ainsi que l'apport de la fusion. Les données parole sont dégradées à deux niveaux différents: -10 dB et 0dB de bruit.

B. Résultats

Le tableau 1 présente les taux d'erreur moyens des systèmes unimodaux (Signature, Parole Indépendante du Texte, Parole Dépendante du Texte) et de fusion (ATC et SVM) sur les 20 différentes bases *BTF*s avec un intervalle de confiance de 95%. Pour atteindre l'erreur totale minimale, nous trouvons que les systèmes ont tendance à rejeter beaucoup d'accès client, car dans notre expérimentation, il y a beaucoup plus d'accès imposteur que d'accès client (12 accès imposteur par rapport à 5 accès client par personne). La dégradation des données de parole à -10dB n'implique pas une augmentation très fort des taux d'erreur, au contrait de ce qui se passe lorsque la parole est dégradée à 0dB (c'est à dire le bruit est aussi fort que le signal de parole), les taux d'erreur augmentent d'un facteur 2.

Nous remarquons que les meilleurs résultats sont obtenus par le système de fusion ATC, mais que les deux systèmes de fusion (ATC et SVM) donnent des résultats similaires lorsque l'on considère l'intervalle de confiance. En effet, dans le cas non bruité, le taux d'erreur moyen baisse de 7,65% (pour le meilleur système unimodal) à 2,21%. Ce qui implique un gain de facteur 3.

TABLEAU 1. L'ERREUR TOTALE MOYENNE DES SYSTÈMES UNIMODAUX ET DES SYSTÈMES DE FUSION AVEC UN INTERVALLE DE CONFIANCE DE 95%

	Model	TE (%)	FA (%)	FR (%)
	Signature	12,34[±0.61]	11,23[±1.22]	14,93[±2.09]
Parole non bruitée	Parole IT	7,65[±0.41]	2,34[±0.32]	20,26[±1.34]
	Parole DT	10,92[±0.55]	5,96[±0.80]	22,60[±2.43]
	ATC	2,21[±0.28]	1,91[±0.41]	4,01[±1.24]
	SVM	2,89[±0.33]	1,75[±0.33]	5,56[±1.05]
Parole -10dB de bruit	Parole IT	9,26[±0.53]	3,78[±0.85]	22,25[±1.64]
	Parole DT	10,78[±0.41]	5,67[±0.61]	22,85[±2.01]
	ATC	2,34[±0.34]	1,70[±0.24]	4,72[±0.95]
	SVM	2,74[±0.31]	1,36[±0.31]	5,99[±1.17]
Parole 0dB De bruit	Parole IT	22,08[±0.79]	7,28[±1.74]	57,17[±3.26]
	Parole DT	17,71[±0.55]	8,62[±1.39]	39,15[±2.93]
	ATC	5,47[±0.61]	3,77[±0.69]	13,32[±2.67]
	SVM	6,69[±0.34]	3,64[±0.78]	13,88[±2.19]

VI. Conclusion et Perspectives

Dans cet article, nous présentons des idées générales d'un système de vérification biométrique. La base BIOMET contenant plusieurs modalités biométriques est brièvement présentée. Deux modalités (Signature en ligne et Parole) sont utilisées pour les premières expériences de vérification. Nous avons construit 3 systèmes de vérification (deux de Parole et un de Signature) et fusionné ces 3 systèmes avec deux méthodes différentes (ATC et SVM). Le signal de parole, qui est souvent bruité par l'environnement, est dégradé à deux niveaux différents pour tester la robustesse des systèmes de vérification de locuteur ainsi que celle des systèmes de fusion. Nous avons vérifié que la fusion améliore beaucoup la performance des systèmes de vérification biométriques unimodaux. Dans tous les cas, le taux d'erreur total baisse de la moitié au moins, en comparaison avec le meilleur système unimodal. Nous obtenons un taux d'erreur total du système de fusion à 2,21% dans le meilleur des cas.

Dans l'avenir, nous allons exploiter d'autres modalités de la base BIOMET comme l'empreinte digitale, le visage... Nous chercherons aussi d'autres méthodes de fusion efficaces, pour rendre plus robustes nos systèmes de vérification.

RÉFÉRENCES

[1] S. Ben-Yacoub, "Multi-Modal Data Fusion for Person Authentification using SVM", *IDIAP Research Report 98-07*, 1998.

[2] S. Ben-Yacoub, Y. Abdeljaoued and E. Mayoraz, “Fusion of Face and Speech Data for Person Identity Verification”, *IEEE Trans. On Neural Networks*, Vol. 10, No 5, 1999, pp. 1065-1074.

[3] B. Gutschoven, P. Verlinde, “Multimodal Identity Verification using Support Vector Machine”, *Fusion 2000*, 2000.

[4] A. Ross, A. Jain and J-Z. Qian, “Information Fusion in Biometrics”, *3rd Int'l Conference on Audio- and Video-Based Person Authentication*, AVBPA, pp. 354-359, Sweden, June 6-8, 2001.

[5] H. Drucker and C. Cortes. “Boosting decision trees”. *Advances in Neural Information Processing Systems*, volume 8. 1996.

[6] C. Sanderson, K. K. Paliwal, “Information Fusion and Person Verification using Speech and Face Information”, *IDIAP Research Report, 02-33*, September 2002.

[7] M. Fuentes, S. Garcia-Salicetti, B. Dorizzi "On-line Signature Verification: Fusion of a Hidden Markov Model and a Neural Network via a Support Vector Machine", *IWFHR8*, August 2002.

[8] P. Verlinde, “A Contribution to Multimodal Identity Verification Using Decision Fusion”, *Ph.D. Thesis*, Department of Signal and Image Processing, Telecom Paris, France, 1999.

[9] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. Leroux-Les Jardins, J. Lunter, Y. Ni, D. Petrovska-Delacretaz, "BIOMET: a Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities", *4th International Conference on Audio and Vidio-Based Biometric Person Authentication*, 2003.

[10] L. Rabiner, B.H. Juang, "Fundamentals of Speech Recognition", *Prentice Hall Signal Processing Series*, 1993.

[11] Furui S., “Cepstral Analysis Technique for Automatic Speaker Verification”, *IEEE Trans. Acoustic, Speech, Signal Processing*, Vol ASSP – 29, 254-272, 1981.

[12] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, Vol. 10, No. 1, pp. 19-41, Jan. 2000.

[13] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, “Classification and Regression Trees”, *Belmont*, CA: Wadsworth, 1984.

[14] Y. Freund, “Boosting a Weak Learning Algorithm by Majority”, *Proceedings of the Third Workshop on Computational Learning Theory*, Morgan-Kaufman, 202-216, 1990.

[15] J.R. Quinlan, "Bagging, Boosting, and C4.5", *Proceedings of the 13th National Conference on Artificial Intelligence*, pp. 725-730, 1996.

[16] J.G.A. Dolfing, “Handwriting Recognition and Verification, a Hidden Markov Approach”, *Ph.D. Thesis*, Philips Electronics N.V., 1998.

[17] L. Breiman, "Bagging predictors", *Machine Learning*, 24(2), pp. 123-140, 1996.

[18] A. Rosenberg, J. DeLong, C-H. Lee, B-H. Juang, and F. Soong. "The Use of Cohort Normalized Scores for Speaker Verification" *International Conference on Spoken Language Processing in Banff*, University of Alberta, 599 - 602, 1992

[19] D. A. Reynolds & Rose R. C. “Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models”, *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72 – 83, 1995.

[20] J. Friedman, T. Hastie and R. Tibshirani, “Additive logistic regression: a statistical view of boosting”, *Dept. of Statistics, Stanford University Technical Report, 1998.*

[21] V. Vapnik, “The Nature of Statistical Learning Theory”, *Statistics for Engineering and Information Science*, Second Edition, Springer, 1999.

[22] A. Papoulis, Probability and Statistics, Prentice Hall, 1990.