

Contributions à la dynamique de frappe au clavier

multibiométrie, biométrie douce et mise à jour de la
référence

Romain GIOT

Sous la direction de :

Christophe ROSENBERGER (ENSICAEN)

Bernadette DORIZZI (Télécom SudParis)

ENSICAEN - Université de Caen - GREYC - CNRS, UMR 6072
École doctorale SIMEM
ANR ASAP

23 octobre 2012

Contexte de la thèse

Usage du numérique par les français

- 78% ont un ordinateur *(CREDOC, 2011)*
- 85% ont un mobile

Comportement

- 15 h/semaine sur internet (17 pour télévision) *(CREDOC, 2010)*
- 77% des internautes utilisent un réseau social *(Médiamétrie, 2012)*

**Il est nécessaire de disposer d'une solution de contrôle
d'accès logique sécurisée**

Nous voulons un système d'authentification

- À bas coût
- Non intrusif
- Permanent
- Faiblement contraint

Méthodes actuelles

Authentification basée sur :

- la **connaissance** (mot de passe)
 - Problèmes : politiques de mots de passe complexes, ...
- la **possession** (token)
 - Problèmes : que faire en cas de vol ?
- ce **qu'on est** (biométrie)
 - ⇒ Seule méthode qui garantissant le lien avec l'individu

Nous voulons un système d'authentification

- À bas coût
- Non intrusif
- Permanent
- Faiblement contraint

Méthodes actuelles

Authentification basée sur :

- la **connaissance** (mot de passe)
 - Problèmes : politiques de mots de passe complexes, ...
- la **possession** (token)
 - Problèmes : que faire en cas de vol ?
- ce **qu'on est** (biométrie)
 - ⇒ Seule méthode qui garantissant le lien avec l'individu

Fonctionnement d'un système biométrique

Enregistrement

- 1 **Acquisition** de donnée(s) biométrique
 - Collecte de la donnée biométrique
 - Extraction des caractéristiques
- 2 Calcul de la **référence biométrique**

Vérification (comparaison 1:1)

- 1 **Acquisition** d'une donnée biométrique (**requête**)
- 2 Calcul d'un **score** de comparaison à la référence
- 3 **Décision** d'acceptation ou de rejet

Fonctionnement d'un système biométrique

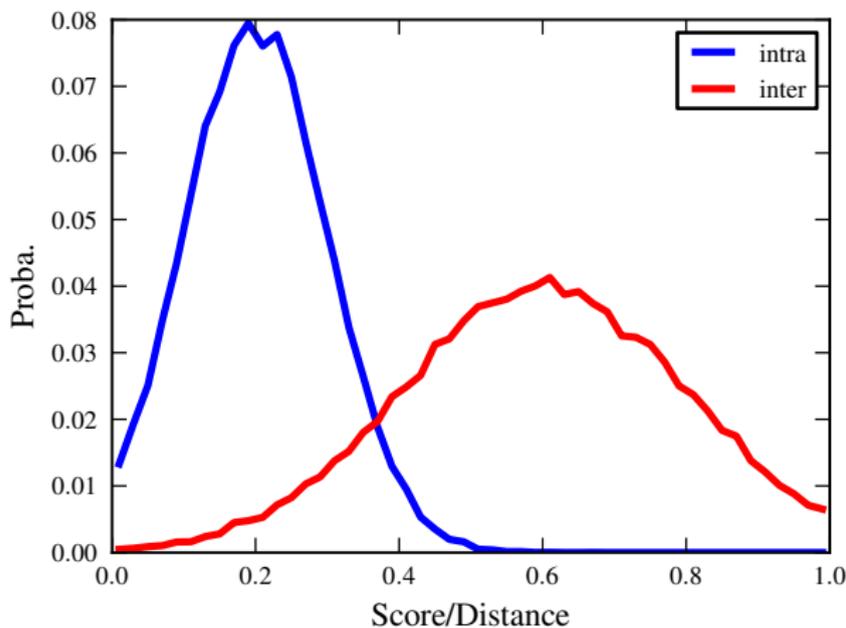
Enregistrement

- 1 **Acquisition** de donnée(s) biométrique
 - Collecte de la donnée biométrique
 - Extraction des caractéristiques
- 2 Calcul de la **référence biométrique**

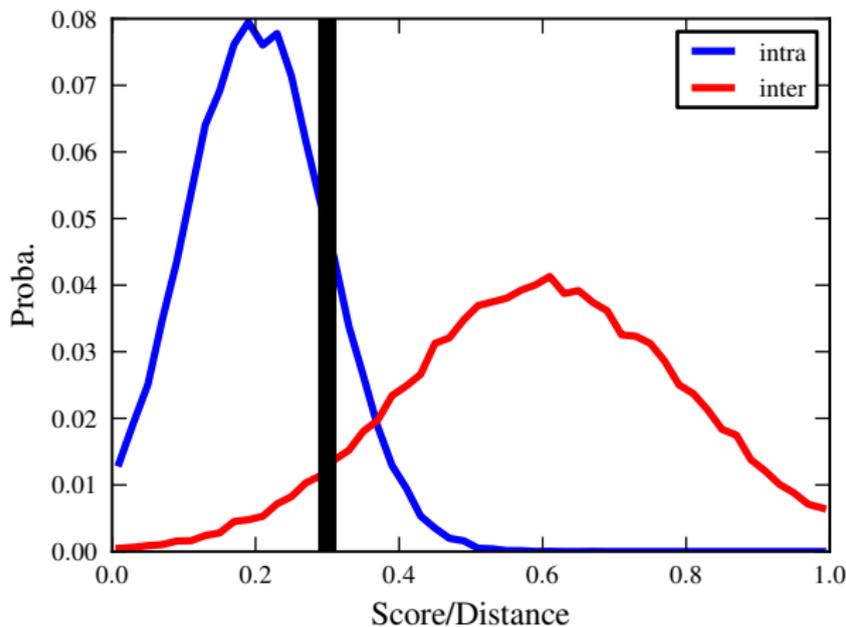
Vérification (comparaison 1:1)

- 1 **Acquisition** d'une donnée biométrique (**requête**)
- 2 Calcul d'un **score** de comparaison à la référence
- 3 **Décision** d'acceptation ou de rejet

Évaluation d'un système biométrique

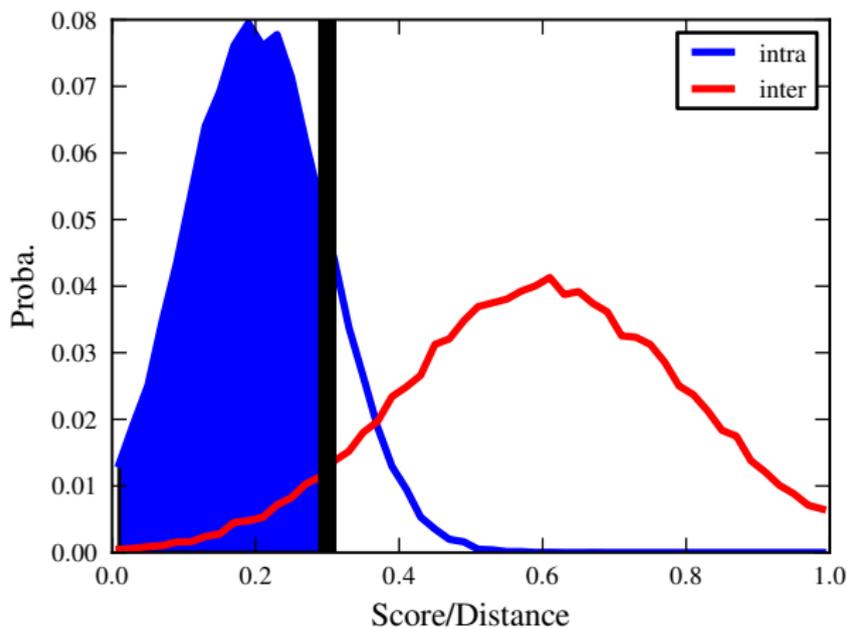


Évaluation d'un système biométrique

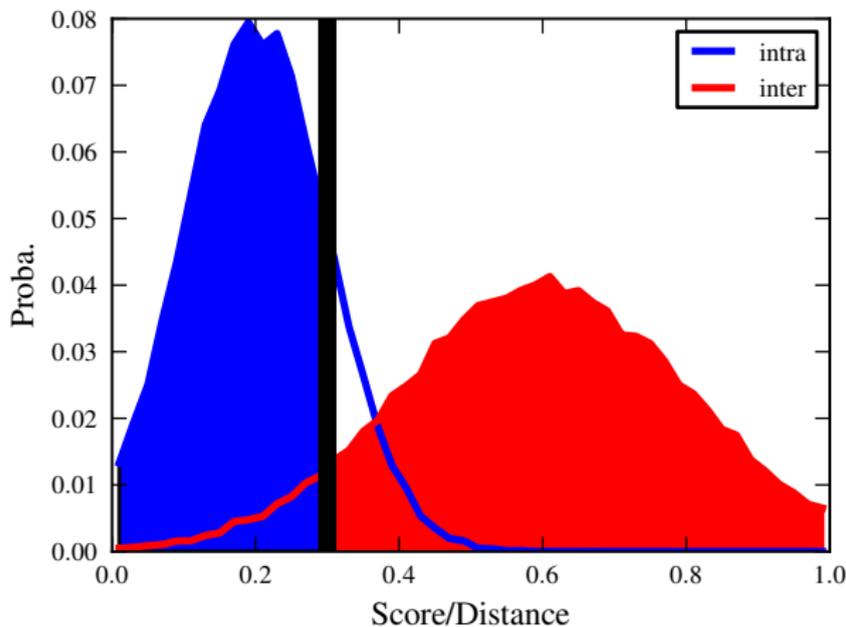


Seuil de décision d'authentification

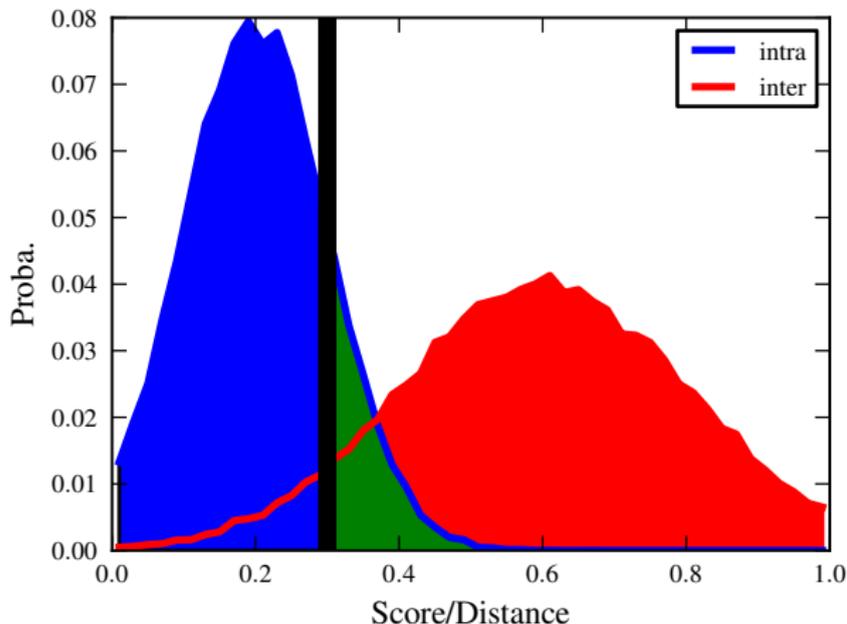
Évaluation d'un système biométrique



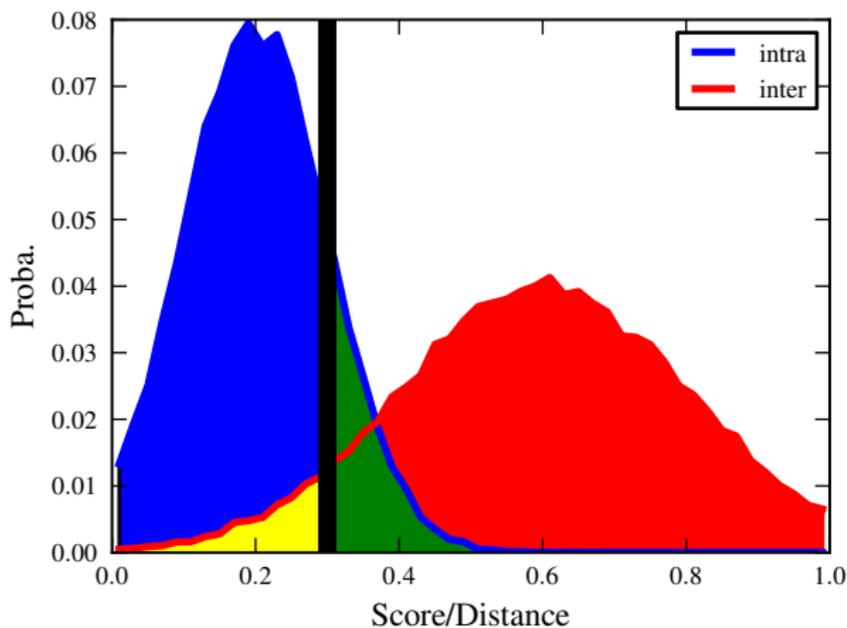
Évaluation d'un système biométrique



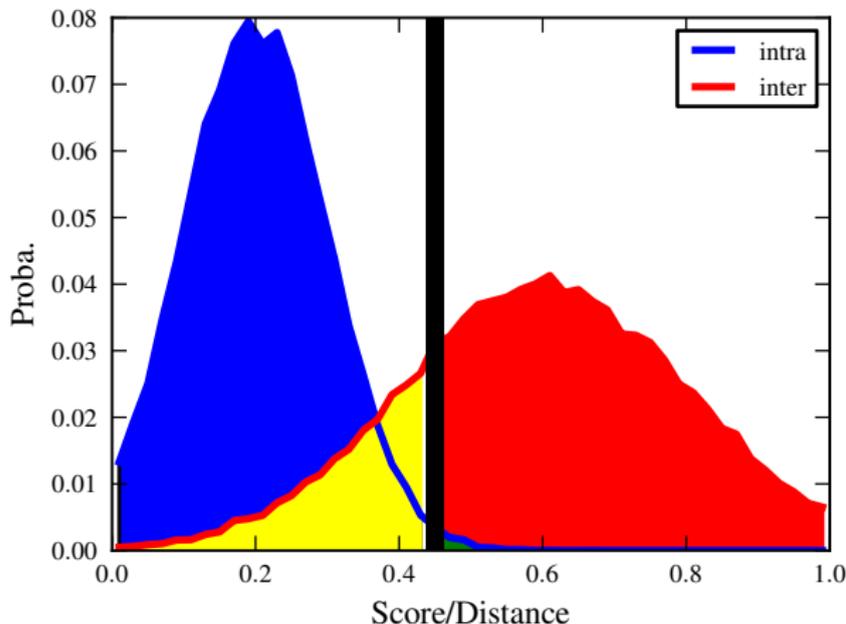
Évaluation d'un système biométrique



Évaluation d'un système biométrique

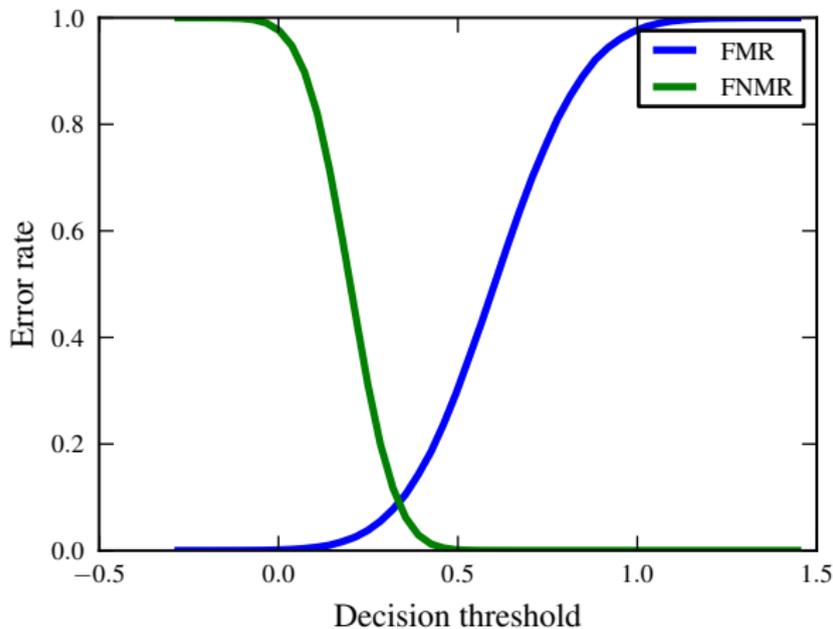


Évaluation d'un système biométrique

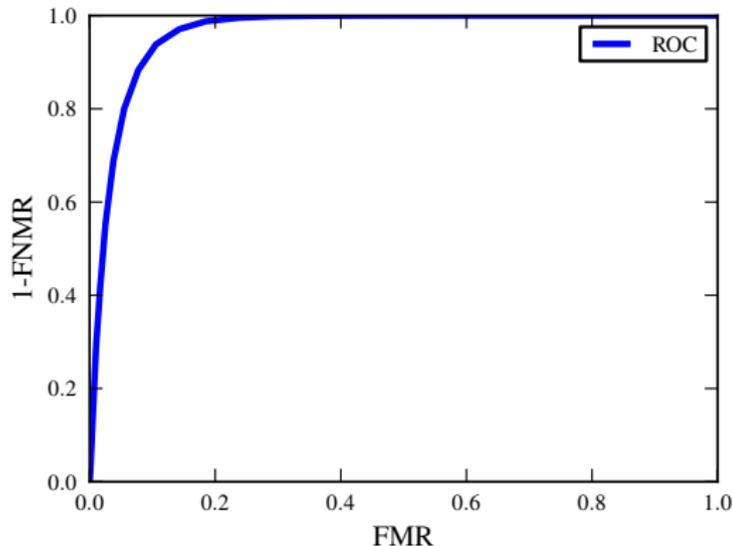


Un seuil différent donne des erreurs différentes

Évaluation d'un système biométrique

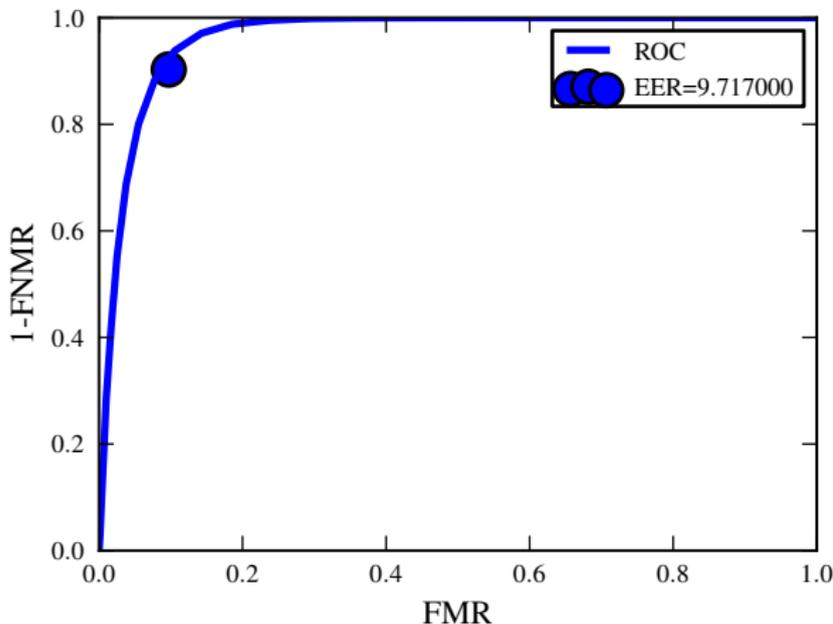


Évaluation d'un système biométrique



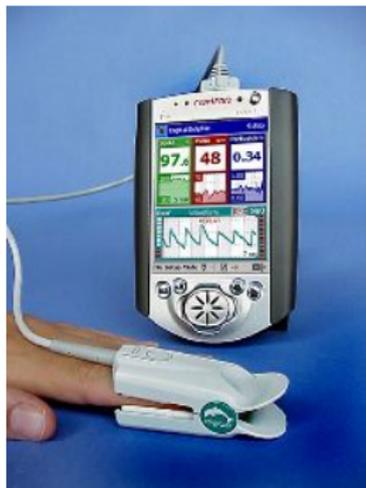
La courbe ROC donne une vision globale des performances de reconnaissance

Évaluation d'un système biométrique



EER/TEE (point de fonctionnement classique)

Modalités biométriques



Biologique



Morphologique



Comportementale

Dynamique De Frappe au clavier

- **Authentification** suivant la façon de **taper au clavier**
- Bien adaptée au contrôle d'**accès logique**

Avantages de cette modalité

- **Peu coûteuse**
- **Non intrusive** / facilement acceptée

Mode d'utilisation

- **Authentification statique**
 - **texte fixe**
- **Authentification continue**
 - **texte libre**

Données utilisées

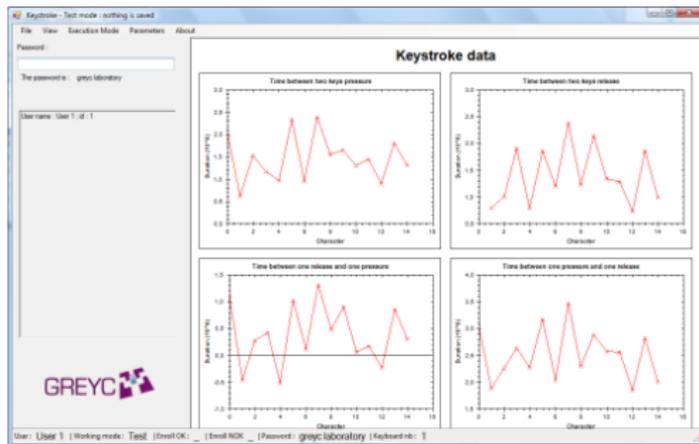
Données brutes

- **Heure** de **pression** des touches
- **Heure** de **relâchement** des touches
- **Code** de la touche concernée

Données extraites

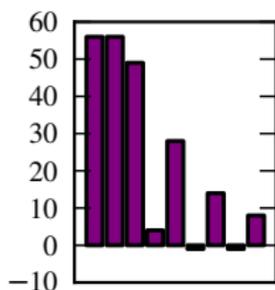
- **1 Pression**
PR Durée de pression d'une touche
- **3 Latences**
PP Durée entre **deux pressions** successives
RR Durée entre **deux relâchements** successifs
RP Durée entre le **relâchement** d'une touche et la **pression** de la suivante

GREYC Keystroke

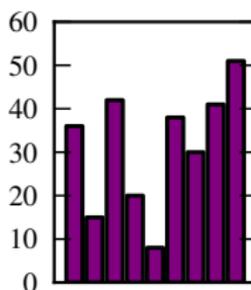


- Visualisation des données extraites
 - Démonstration du fonctionnement de la DDF
 - **Création d'un des jeux de données de l'état de l'art**
- (Giot et al., 2009)

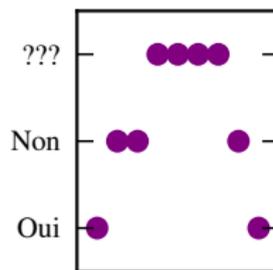
Limites des travaux de la littérature



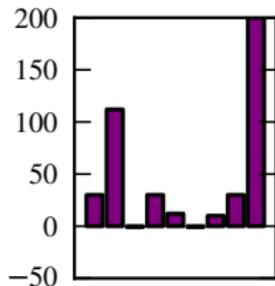
Durée



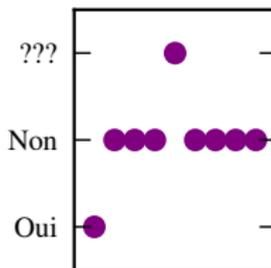
Nb individus



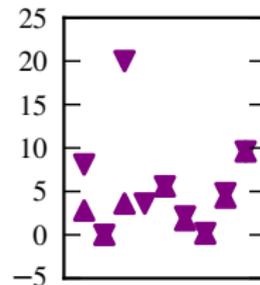
Acquis. contrôlée



Nb ex./enr.



Seuil global



FMR/FNMR

Limites de la Dynamique De Frappe

- Dépend de l'**état émotionnel** de l'individu *(Epp et al., 2011)*
- Nécessite **beaucoup** de données pour l'**enregistrement**
- Les **performances** sont instables au cours du temps

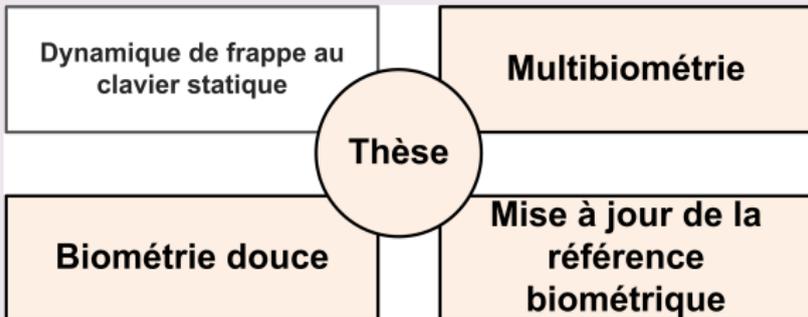
(Giot et al., 2011)

Objectifs de la thèse

Améliorer la Dynamique de Frappe en terme de

- **sécurité** / diminution du EER
- **usage** / utilisation de peu de captures

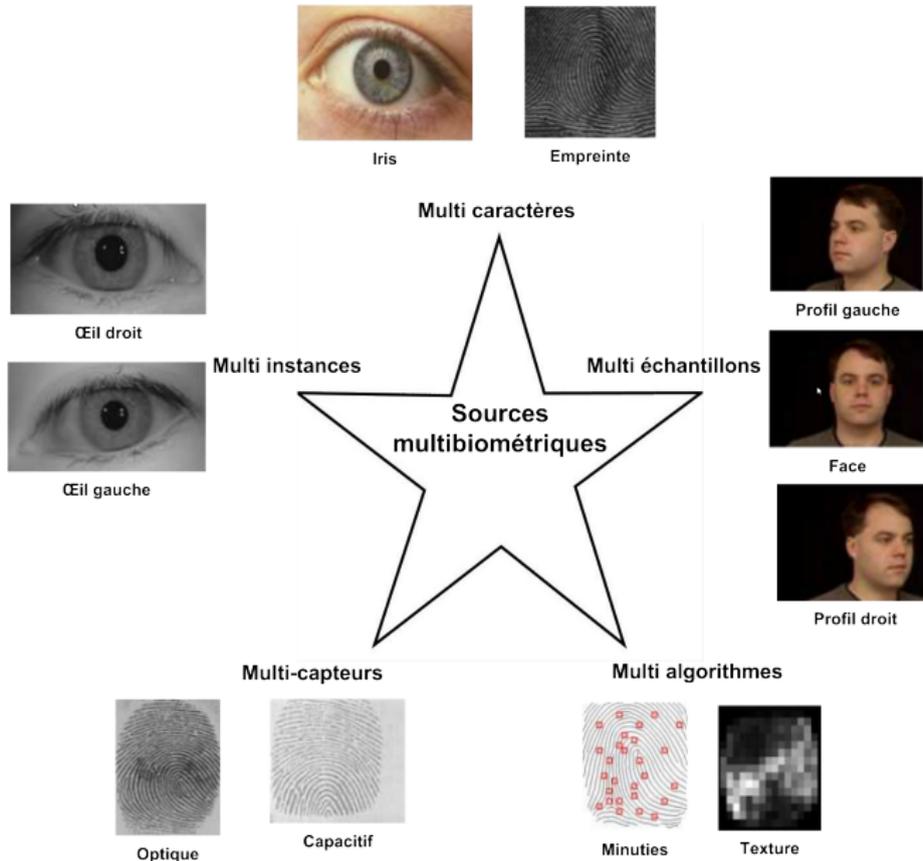
Contribution



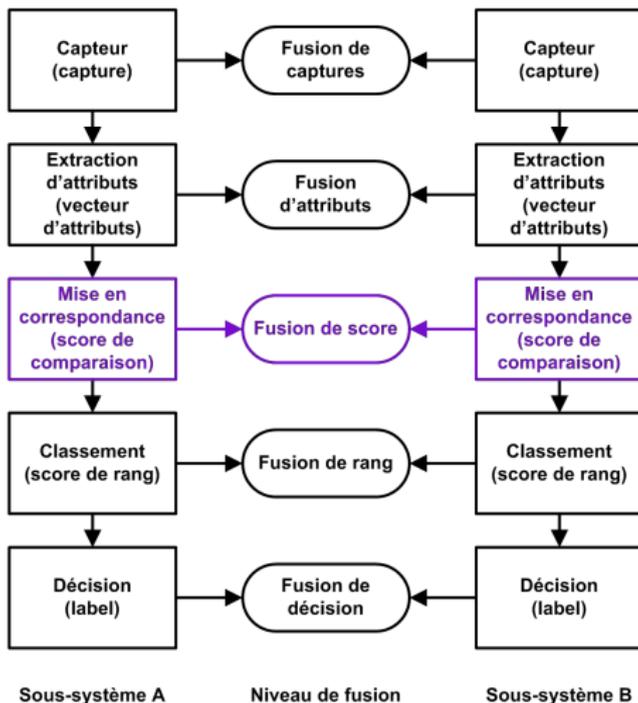
Sommaire

- 1 Positionnement du problème
- 2 Multibiométrie**
- 3 Biométrie douce
- 4 Mise à jour de la référence biométrique

Qu'est-ce que la multimodalité ?



Niveaux de fusion



État de l'art succinct

Beaucoup de travaux dans la littérature

(Ross et al., 2006)

Fusion de score classique

- Règle somme : $sum = \sum_{i=1}^C s_i$ (Kittler, 1998)
- Règle produit : $mul = \prod_{i=1}^C s_i$
- Règle minimum/maximum : $min = \min_{i=1}^C s_i$
- **Somme pondérée** : $ga1 = \sum_{i=1}^C w_i * s_i$

Objectif : système multibiométrique peu coûteux

Modalités sélectionnées

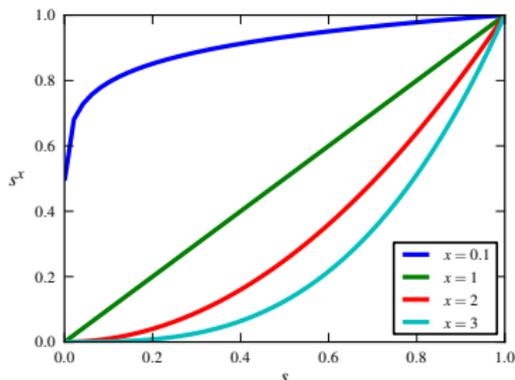
- Dynamique De Frappe au clavier (DDF)
- Reconnaissance Faciale (RF)

Justifications

- **Pas de matériel supplémentaire**
 - Clavier et webcam présents sur tous les ordinateurs
- **Peu contraignant**
 - Captures collectées à la saisie du mot de passe

Fonctions de fusion définies a priori

- $ga2 = \prod_{i=1}^c s_i^{x_i}$
- $ga3 = \sum_{i=1}^c w_i * s_i^{x_i}$



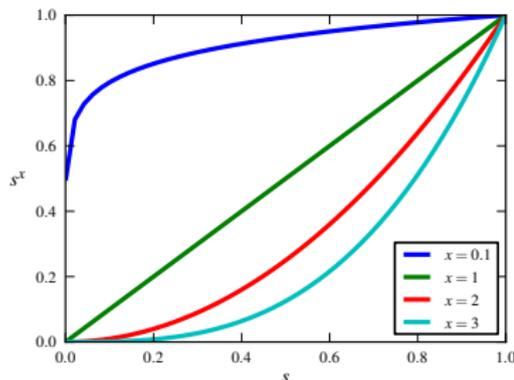
Paramétrage par algorithme génétique

- Méthode évolutionnaire
- génotype = **paramétrage** de la fonction (w et x)
- Évaluation = EER sur base d'apprentissage

(Holland, 1975)

Fonctions de fusion définies a priori

- $ga2 = \prod_{i=1}^c s_i^{x_i}$
- $ga3 = \sum_{i=1}^c w_i * s_i^{x_i}$



Paramétrage par algorithme génétique

- Méthode évolutionnaire
- génotype = **paramétrage** de la fonction (**w** et **x**)
- Évaluation = EER sur base d'apprentissage

(Holland, 1975)

Fonctions de fusions définies a posteriori

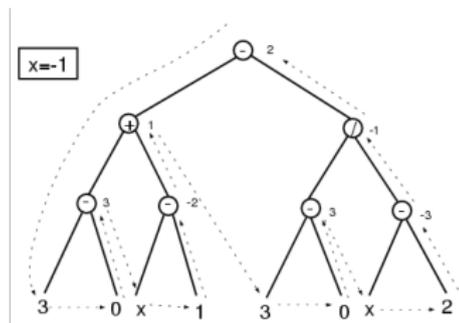
Nous voulons trouver une **fonction** de fusion **optimale**

Programmation génétique *(Koza, 1990)*

- Méthode évolutionnaire
- génotype = 1 **fonction** de fusion
- Évaluation = EER sur base d'apprentissage

Paramétrage

- Nœuds : +, -, *, /, min, max, avg
- Feuilles : scores, constantes



(Koza, 1994)

Définition d'une base chimérique

Bases initiales

- RF : **AR** *(Martinez et al., 1998)*
 - 120 individus
 - 26 exemples par utilisateurs
- DDF : **GREYC** *(Giot et al., 2009)*
 - 100 individus
 - 60 exemples par utilisateur



Base chimérique obtenue

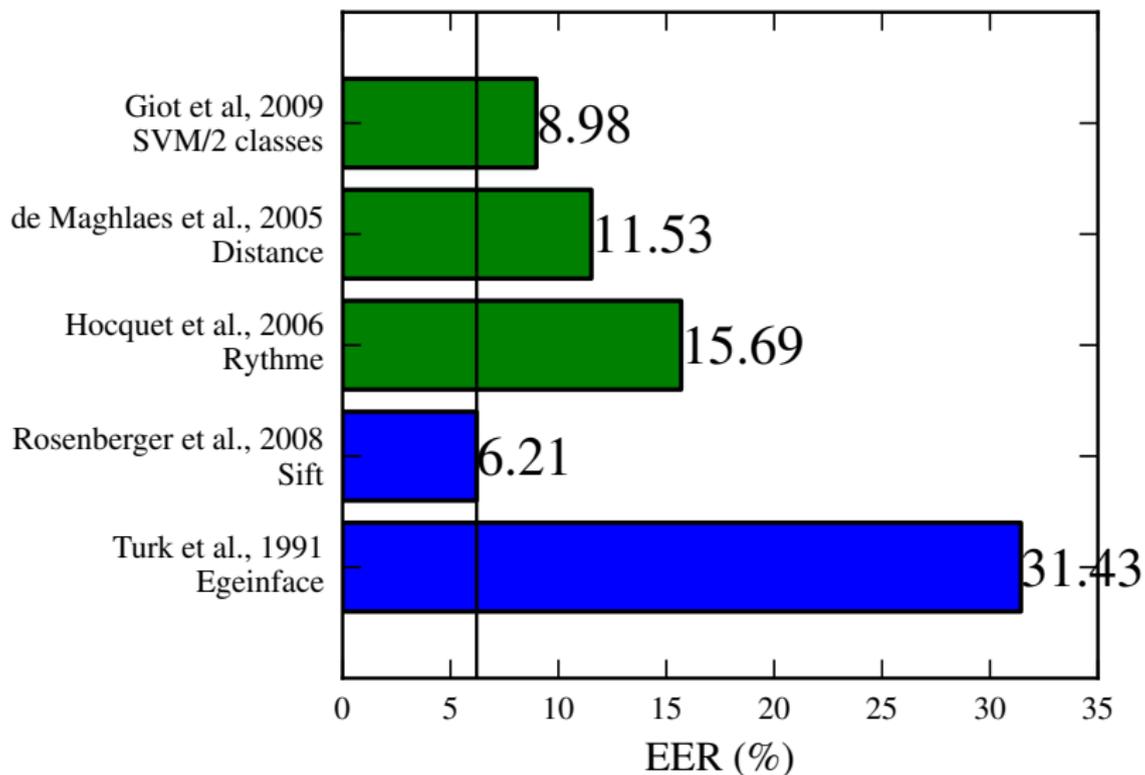
- 100 individus
- 26 exemples par utilisateurs
- utilisateur i DDF = utilisateur i RF

Protocole expérimental

Configuration des exemples	10/enregistrement 16/vérification
Nombre de n-uplets de scores	intra= $100 \times 16 = 1600$ inter = $100 \times 16 \times 99 = 158400$
Séparation des données	50% apprentissage 50% validation
Systèmes monomodaux	3 méthodes de DDF 2 méthodes de RF
Méthodes de fusion	sum/min/mul/somme pondérée/SVM
Normalisation des scores	tanh

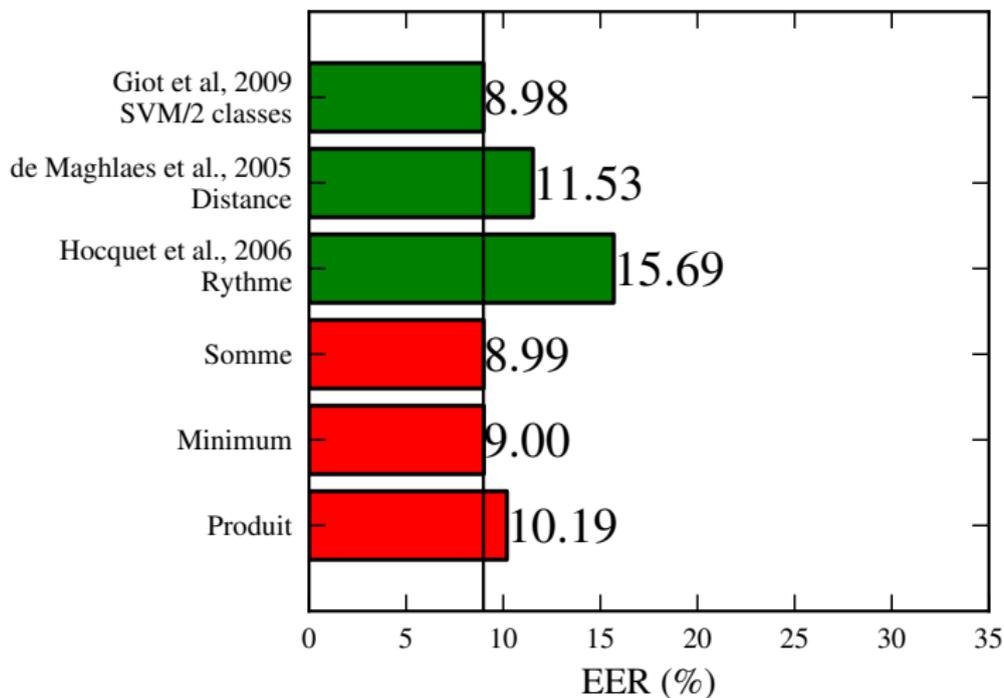
(Jain et al., 2005)

Performances des systèmes mono-modaux



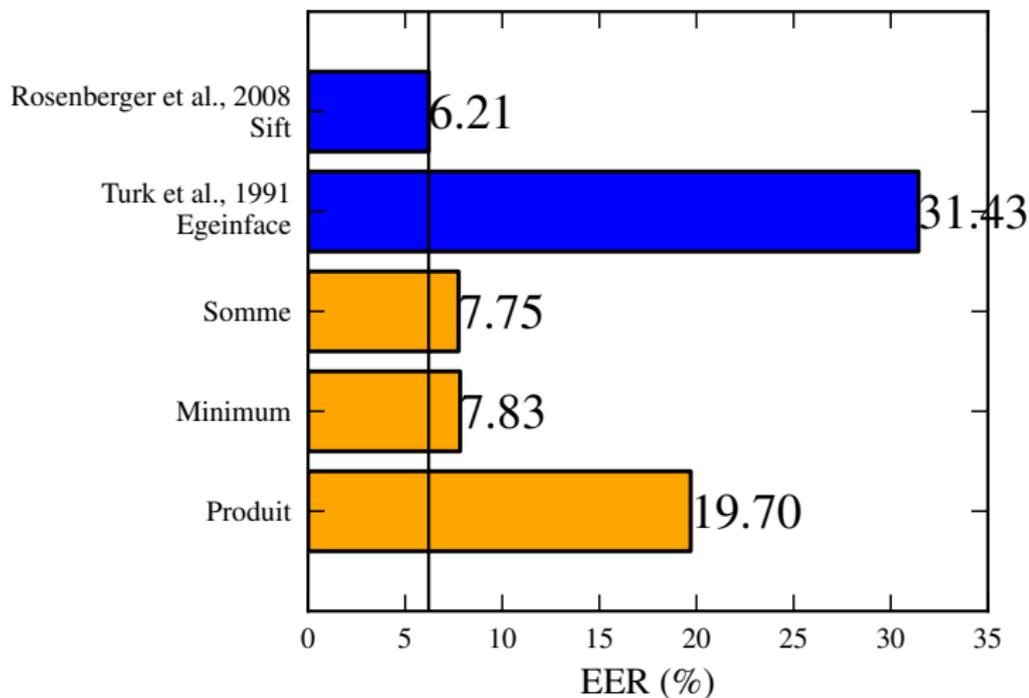
Les performances varient d'un système à l'autre

Fusion Dynamique De Frappe uniquement



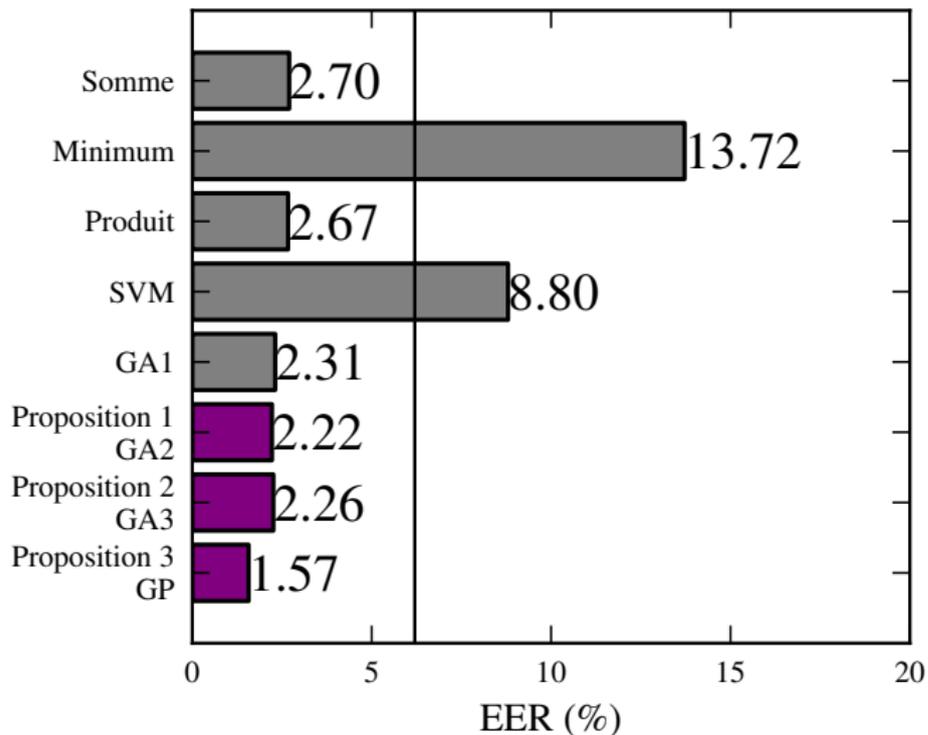
Fusionner tous les systèmes de DDF n'apporte pas de gain

Fusion Reconnaissance Faciale uniquement



Fusionner tous les systèmes de RF n'apporte pas de gain

État de l'art et propositions



Fusion de tous les systèmes monomodaux

Discussion

Résultats

- Important de fusionner des systèmes complémentaires
⇒ **DDF** et **RF** fortement **complémentaires**
- Mise en évidence de l'**intérêt** de nos **contributions**

Limites

- Analyse de l'**intervalle** de confiance à effectuer
- **Complexité** de calcul de la **programmation génétique**

Sommaire

- 1 Positionnement du problème
- 2 Multibiométrie
- 3 Biométrie douce**
- 4 Mise à jour de la référence biométrique

Définition de la biométrie douce

Domaine **nouveau**, **sans consensus** sur la définition.

(Jain et al., 2004; Dantcheva et al., 2011)

Définition courante

Un **caractère** fournissant de l'information sur l'individu, mais **manquant** d'**unicité** et de **permanence** pour **différencier** suffisamment deux individus.

(Jain et al., 2004)

Données et sources exploitables

- Différents types d'informations
 - Genre, âge, ethnie, présence de lunettes, ...
- Information **connue** à priori (passeport, ...)
- Information **extraite** automatiquement (photo, ...)

Définition de la biométrie douce

Domaine **nouveau**, **sans consensus** sur la définition.

(Jain et al., 2004; Dantcheva et al., 2011)

Définition courante

Un **caractère** fournissant de l'information sur l'individu, mais **manquant** d'**unicité** et de **permanence** pour **différencier** suffisamment deux individus.

(Jain et al., 2004)

Données et sources exploitables

- Différents types d'informations
 - Genre, âge, ethnie, présence de lunettes, ...
- Information **connue** à priori (passeport, ...)
- Information **extraite** automatiquement (photo, ...)

Motivations pour étudier la biométrie douce

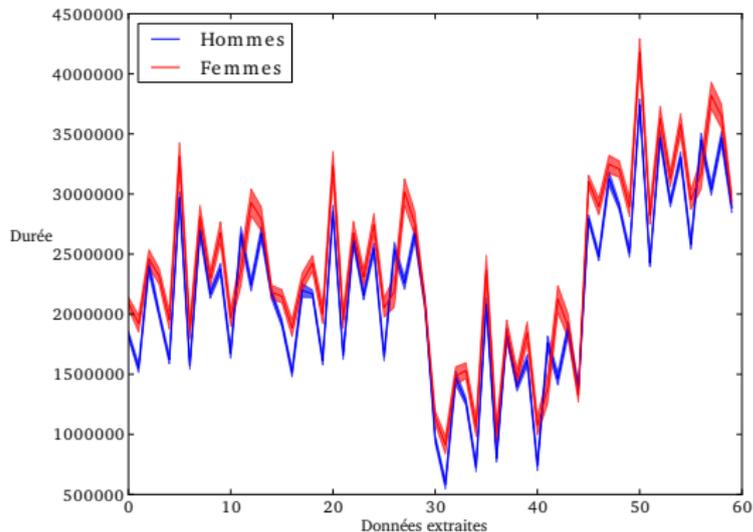
Existant

- **Beaucoup** de travaux en **Reconnaissance Faciale**
 - ethnie, genre, âge, . . . *(Alexandre, 2010; Dantcheva et al., 2011)*
- Mais **peu** en **Dynamique De Frappe**
 - état émotionnel *(Epp et al., 2011)*
 - classe de configuration *(Hocquet et al., 2007)*

Idée

- Les performances en DDF ne sont pas importantes
 - ⇒ il est intéressant de **classer** les utilisateurs, plutôt que les authentifier
 - ⇒ il est intéressant d'**utiliser** la biométrie douce pour **augmenter** les performances de **reconnaissance**

Les hommes et les femmes tapent différemment



Intervalle de confiance à 95% des données extraites d'un mot de passe particulier pour les hommes et les femmes.

Méthode de reconnaissance de genre

Données utilisées

- Vecteur de temps de DDF
- Étiquette de genre (-1/femmes, +1/hommes)

Classifieur utilisé

- Séparateur à Vaste Marge
- Probabilité d'appartenance à la classe identifiée

(Vapnik, 2000)

(Chang et al., 2001)

Score de reconnaissance du genre

$$\frac{1 - \text{prédiction} * \text{probabilité}}{2} \quad (1)$$

Protocole expérimental

Données extraites

- Évaluation pour chaque type de données extraites
 - PP, RR, RP, PR, V

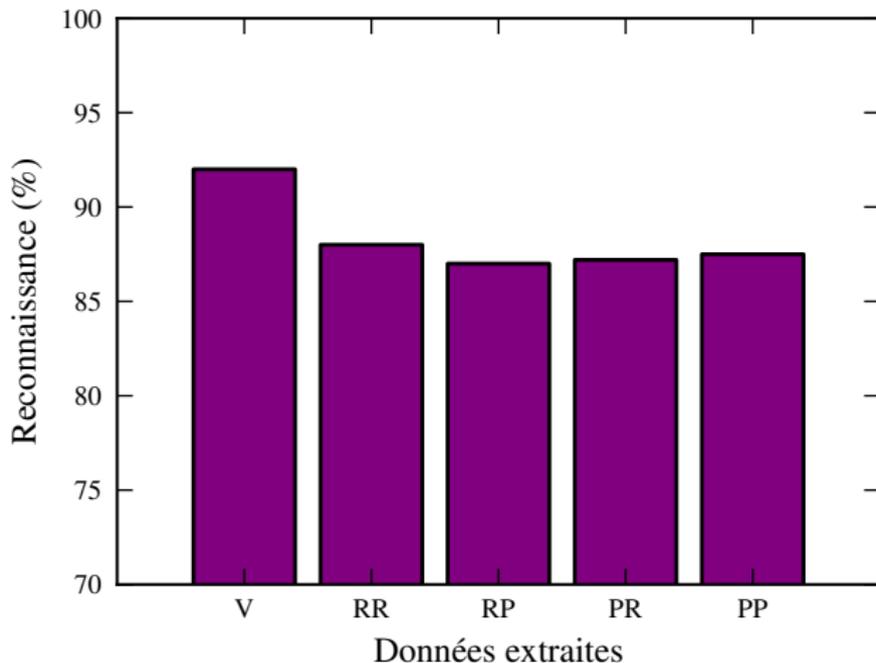
Classifieur

- Données normalisées par zscore, $z = \frac{x-\mu}{\sigma}$
- Validation croisée
- Recherche exhaustive des paramètres du SVM (C, γ)

Jeu de données

- Base GREYC
- 35 hommes/35 femmes conservés

Performance de reconnaissance du genre



Reconnaissance du genre en fonction du type de données extraites

Authentification utilisant le genre comme donnée additionnelle

Données

- Score de reconnaissance **biométrique**
- Score de reconnaissance de **genre**

Fusion biométrie + genre

$$\text{distance_biométrique} + \underbrace{\alpha * \text{abs}(\mu_{\text{genre}} - \text{distance_genre})}_{\text{pénalité}} \quad (2)$$

α configuré empiriquement, μ_{genre} le **score moyen** de genre des données d'enregistrement de l'utilisateur

Authentification utilisant le genre comme donnée additionnelle

Données

- Score de reconnaissance **biométrique**
- Score de reconnaissance de **genre**

Fusion biométrie + genre

$$\text{distance_biométrique} + \underbrace{\alpha * \text{abs}(\mu_{\text{genre}} - \text{distance_genre})}_{\text{pénalité}} \quad (2)$$

α configuré empiriquement, μ_{genre} le **score moyen** de genre des données d'enregistrement de l'utilisateur

Protocole expérimental

Scénarios testés

DESCRIPTION	ENREGISTREMENT	VALIDATION
Pas de genre (scénario de base)	-	-
Étiquetage manuel	Manuel	Manuel
Étiquetage semi-automatique	Manuel	Auto.
Étiquetage automatique	Auto.	Auto.

Paramètres

- Validation croisée à deux ensembles
- Utilisateurs de l'ensemble de validation distincts de celui d'apprentissage

Protocole expérimental

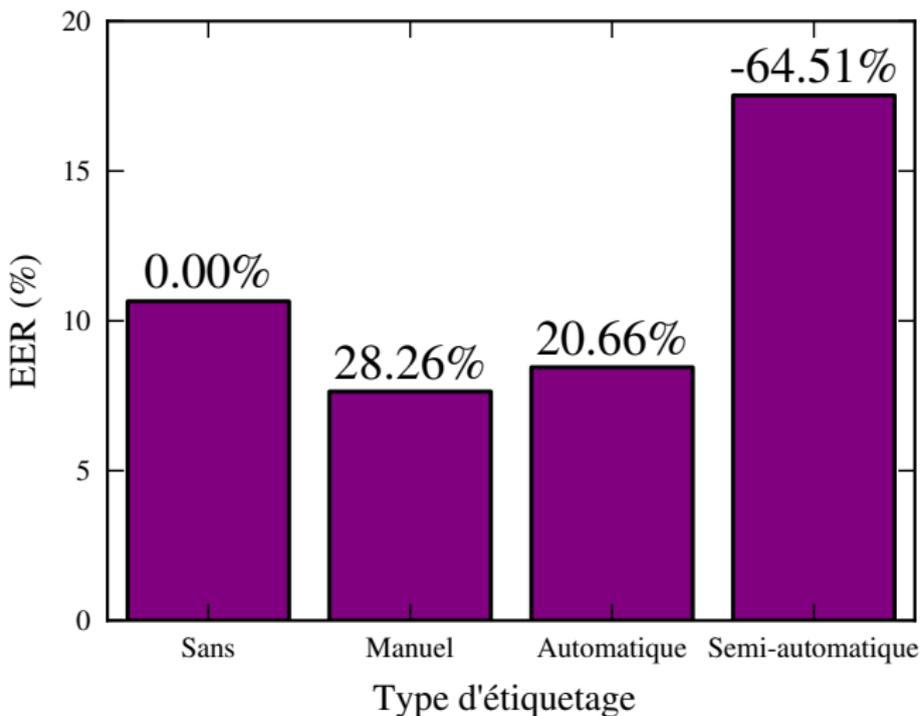
Scénarios testés

DESCRIPTION	ENREGISTREMENT	VALIDATION
Pas de genre (scénario de base)	-	-
Étiquetage manuel	Manuel	Manuel
Étiquetage semi-automatique	Manuel	Auto.
Étiquetage automatique	Auto.	Auto.

Paramètres

- Validation croisée à deux ensembles
- Utilisateurs de l'ensemble de validation distincts de celui d'apprentissage

Performance d'authentification avec genre



Un gain de performance est utilisé en utilisant le genre

Motivation pour l'utilisation du texte libre

Constat

- La reconnaissance du genre sur texte fixe fonctionne à 90%
- MAIS** les cas d'utilisation ne sont pas nombreux

Idée

- ⇒ Il est nécessaire de vérifier le texte libre
- L'âge peut être une catégorie supplémentaire

Méthode de reconnaissance sur texte libre

Données extraites

- Vecteur de taille fixe, indépendant du texte
- Utilisation des digraphes communs dans la langue
- RR, RP, PR, PP

Classifieur utilisé

- Séparateur à Vaste Marge
- Degré de confiance que l'utilisateur soit un homme (ou une personne de moins de 30 ans)

Protocole expérimental

Base publique utilisée

(Bello et al., 2010)

- 55 volontaires
 - 40 hommes/15 femmes
 - 29 moins de 30 ans / 26 plus de 30 ans

- 15 paragraphes

ex : *porque las paredes de la casa y las de la cuadra se han derrumbado encima de todo lo que habia en la casa, sin excluir a los carneros, los gansos y las gallinas.*

Scénarios testés

- Texte fixe/connu
- Texte libre/inconnu
- Plusieurs phrases

Protocole expérimental

Base publique utilisée

(Bello et al., 2010)

- 55 volontaires
 - 40 hommes/15 femmes
 - 29 moins de 30 ans / 26 plus de 30 ans

- 15 paragraphes

ex : *porque las paredes de la casa y las de la cuadra se han derrumbado encima de todo lo que habia en la casa, sin excluir a los carneros, los gansos y las gallinas.*

Scénarios testés

- **Texte fixe**/connu
- **Texte libre**/inconnu
- **Plusieurs** phrases

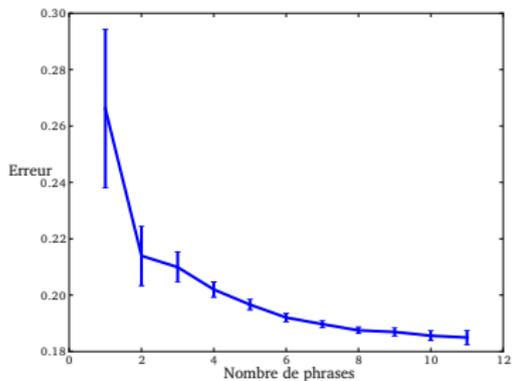
Reconnaissance des catégories grâce à une phrase

Erreur de classification :

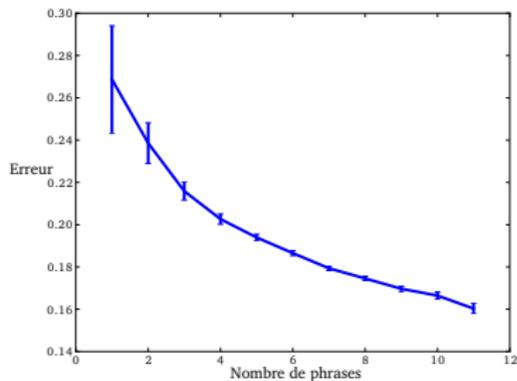
	Texte fixe	Texte libre
Genre	25,5%	26,8%
Âge	25,7%	26,9%

- Performances dépendantes de la phrase
 - Différences entre genre et âge
- **Difficile à comparer avec l'état de l'art**

Résultats en fusionnant plusieurs phrases



Genre



Catégorie d'âge

Discussion

Biométrie douce sur texte fixe

- Reconnaissance du genre possible à 90%
- Permet d'améliorer l'authentification par DDF de 20%

Biométrie douce sur texte libre

- Forte corrélation du genre
- Forte corrélation de catégorie d'âge

Limites

- Nécessité de généraliser les résultats avec d'autres bases plus conséquentes

Sommaire

- 1 Positionnement du problème
- 2 Multibiométrie
- 3 Biométrie douce
- 4 Mise à jour de la référence biométrique**

Motivation à l'utilisation de la mise à jour

Les systèmes font des erreurs de classification

- méthode de vérification, forte variabilité intra-classe, conditions d'acquisitions différentes, capteur différent, manque de coopération, **vieillessement de la donnée biométrique**...

Conséquences

- Gêne pour l'utilisateur
- Faible sécurité du système

La mise à jour de la référence est une solution à ce problème

État de l'art

- La littérature concerne le **visage** et les **empreintes digitales**
 - La mise à jour est majoritairement faite **hors ligne**
 - Une **hiérarchie** existe, mais est **limitée** à ces modalités
- ⇒ **Elle n'est pas applicable directement à la DDF**

Éléments d'un système de mise à jour

- **Périodicité**

En-ligne, hors ligne

- **Mode**

Supervisé,
semi-supervisé

- **Référence**

Unique exemple,
galerie, calcul

- **Stratégie de mise à jour**

Super modèle,
remplacement,
édition, mise à jour
d'un classifieur

- **Critère**

Temps, seuil, qualité,
oracle

Éléments d'un système de mise à jour

- **Périodicité**

En-ligne, hors ligne

- **Mode**

Supervisé,
semi-supervisé

- **Référence**

Unique exemple,
galerie, calcul

- **Stratégie de mise à jour**

Super modèle,
remplacement,
édition, mise à jour
d'un classifieur

- **Critère**

Temps, seuil, qualité,
oracle

Éléments d'un système de mise à jour

- **Périodicité**

En-ligne, hors ligne

- **Mode**

Supervisé,
semi-supervisé

- **Référence**

Unique exemple,
galerie, calcul

- **Stratégie de mise à jour**

Super modèle,
remplacement,
édition, mise à jour
d'un classifieur

- **Critère**

Temps, seuil, qualité,
oracle

Éléments d'un système de mise à jour

- **Périodicité**

En-ligne, hors ligne

- **Mode**

Supervisé,
semi-supervisé

- **Référence**

Unique exemple,
galerie, calcul

- **Stratégie de mise à jour**

Super modèle,
remplacement,
édition, mise à jour
d'un classifieur

- **Critère**

Temps, seuil, qualité,
oracle

Éléments d'un système de mise à jour

- **Périodicité**

En-ligne, hors ligne

- **Mode**

Supervisé,
semi-supervisé

- **Référence**

Unique exemple,
galerie, calcul

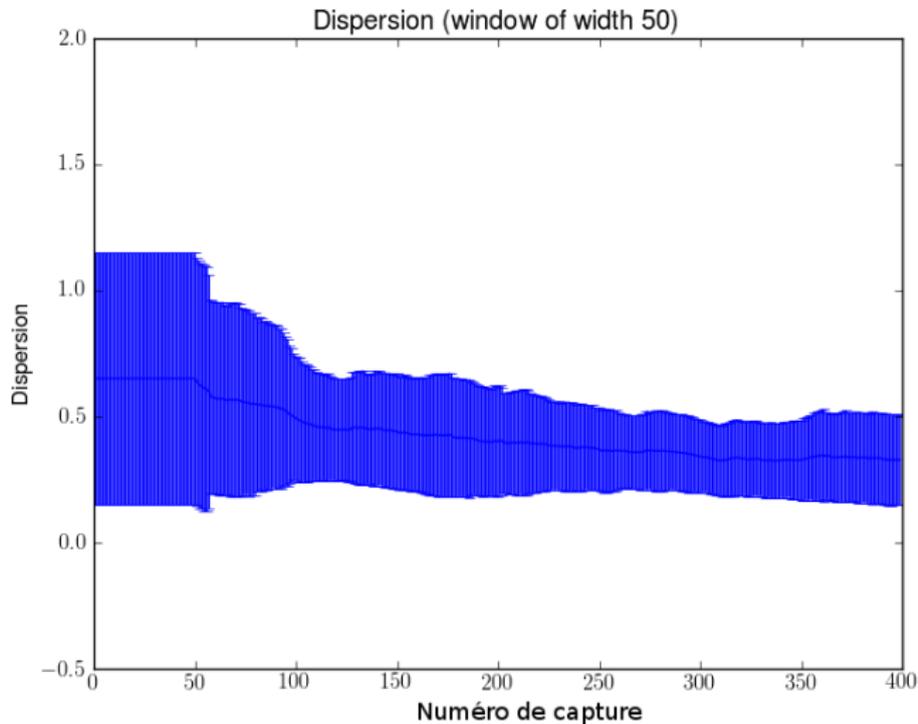
- **Stratégie de mise à jour**

Super modèle,
remplacement,
édition, mise à jour
d'un classifieur

- **Critère**

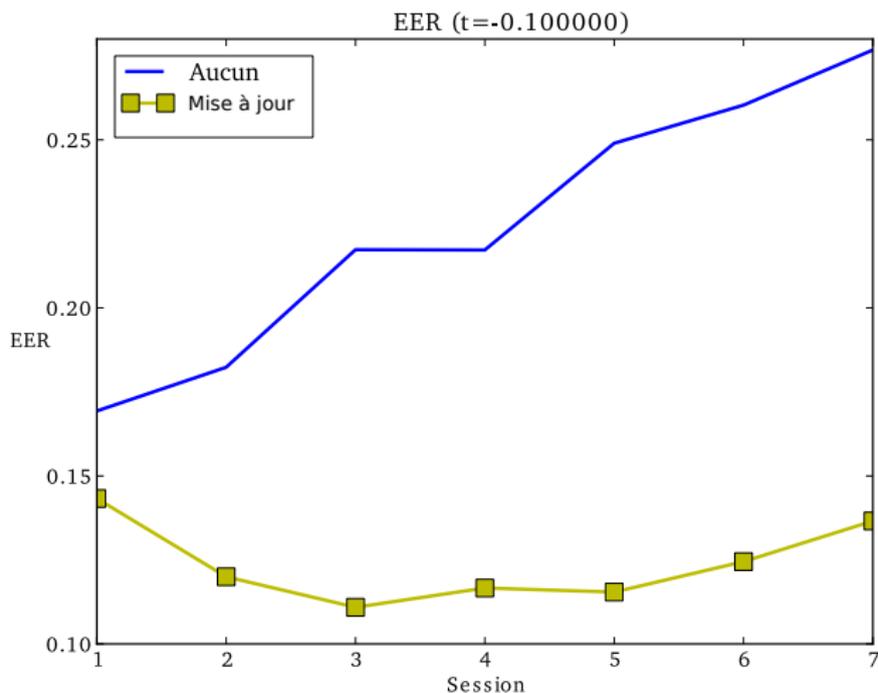
Temps, seuil, qualité,
oracle

Les saisies se stabilisent avec le temps



Moyenne et écart type suivant l'ensemble des utilisateurs de la mesure de dispersion des saisies

Dégradation des performances de reconnaissance au cours du temps



La mise à jour limite la dégradation des performances observées au cours du temps

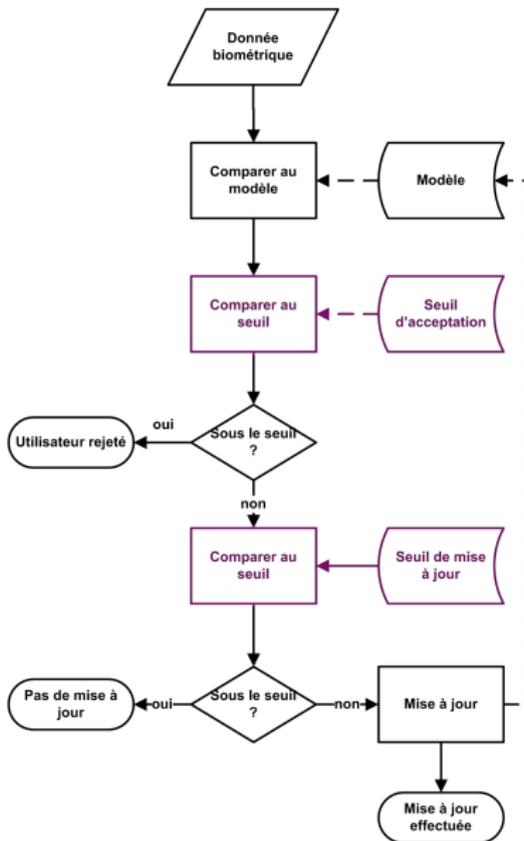
Mise à jour de la référence

- Super modèle
- Méthodes de sélection
- Méthodes de remplacement
 - Fenêtre glissante**
- Méthodes d'ajout
 - Fenêtre croissante**

(Jiang et Ser, 2002)

(Uludag et al., 2004 ; Freni, 2010)

Double seuillage



Auto-apprentissage

- **Étiquetage** automatique de données **non étiquetées**
- **Mise à jour** du classifieur en utilisant les données non étiquetées prédites avec une **forte probabilité**
- **Répétition** de l'étape d'étiquetage

(*Rolli et al., 2006*)

Co-apprentissage

- Même principe que l'auto-apprentissage
- Utilisation de **deux classifieurs**
 - Deux modalités différentes (Rattani et al., 2008)
 - Une seule modalité (Zhao et al., 2011)
- **Mise à jour** d'un classifieur si **forte prédiction** avec l'**autre** classifieur

Définition de notre système hybride

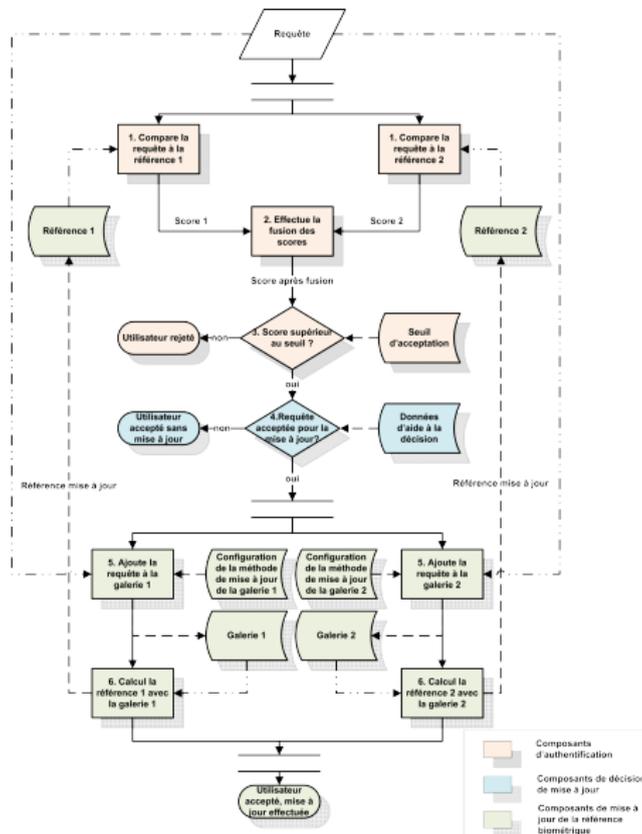
Combinaison

- Auto-apprentissage
- Co-apprentissage

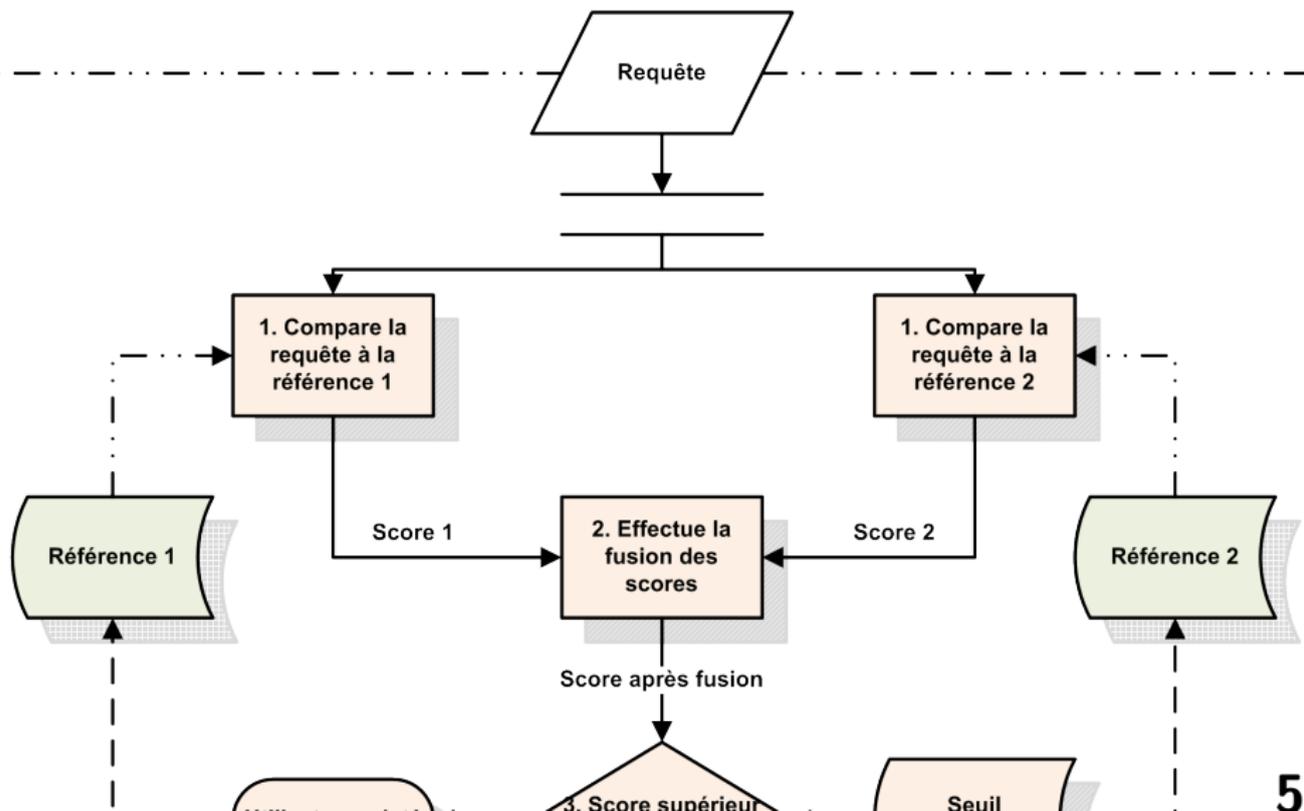
Fonctionnement

- **Une** seule **modalité**
- **Plusieurs références** par utilisateur
- Une méthode de **mise à jour différente** par référence
- **Évolution** en **parallèle** des références

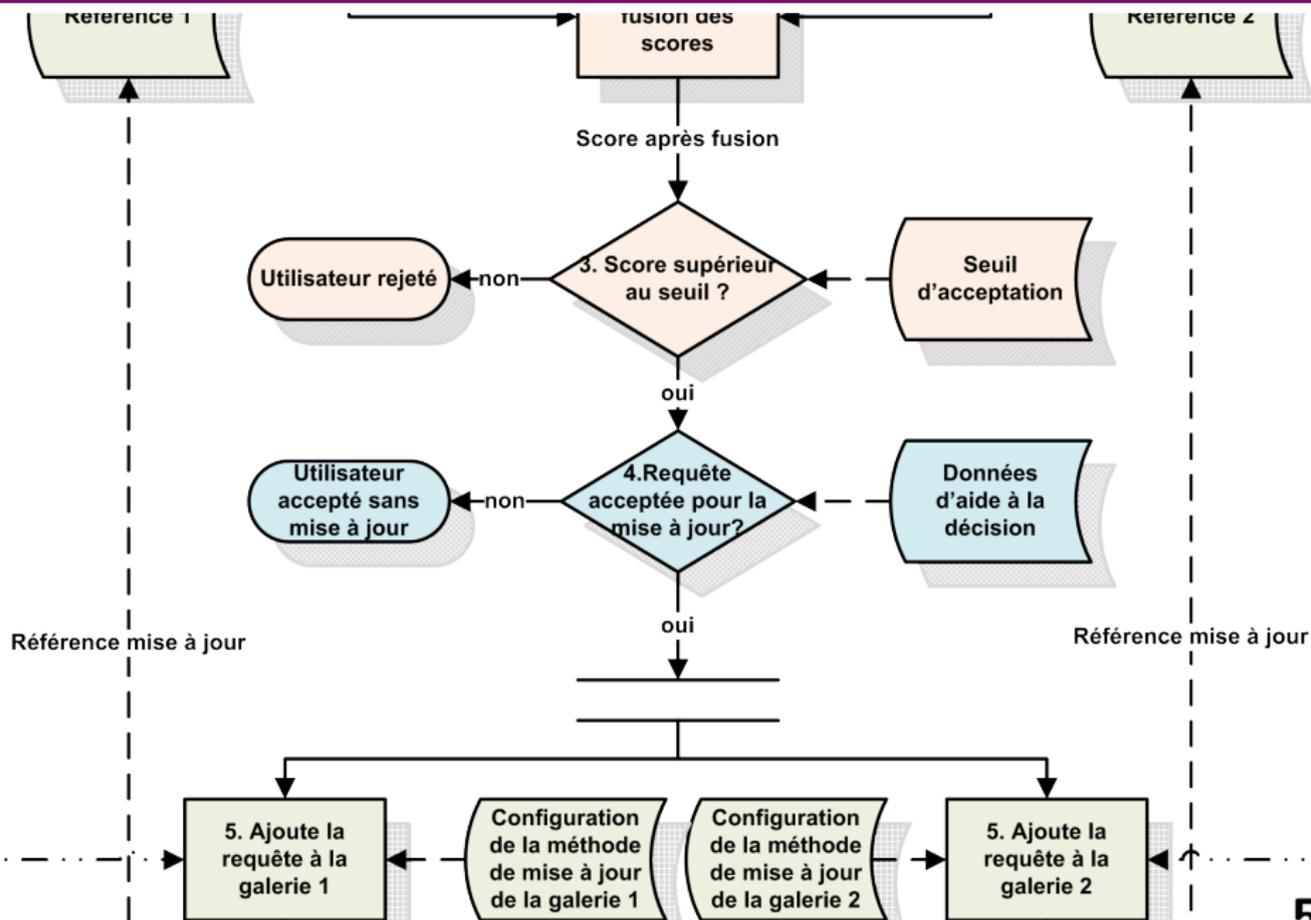
Fonctionnement du système hybride



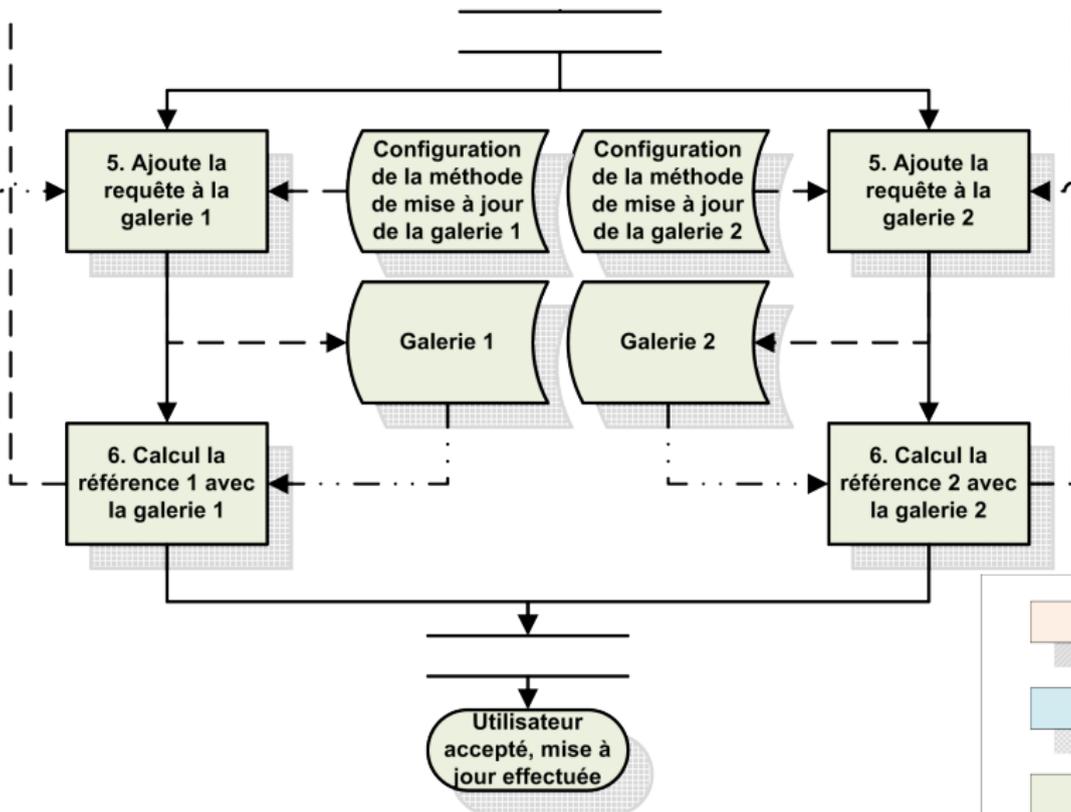
Fonctionnement du système hybride



Fonctionnement du système hybride



Fonctionnement du système hybride



Comparaison du système à l'état de l'art

	Auto-apprentissage	Co-apprentissage	Apprentissage hybride
Une seule modalité	oui	non	oui
Plusieurs types de classifieurs	non	oui (un par modalité)	choix d'implémentation (un par sous-referencé ou le même pour tous)
Source de décision de mise à jour	score du classifieur	désaccord entre les deux classifieurs	Score agrégé à partir des scores produits par la comparaison aux sous-références

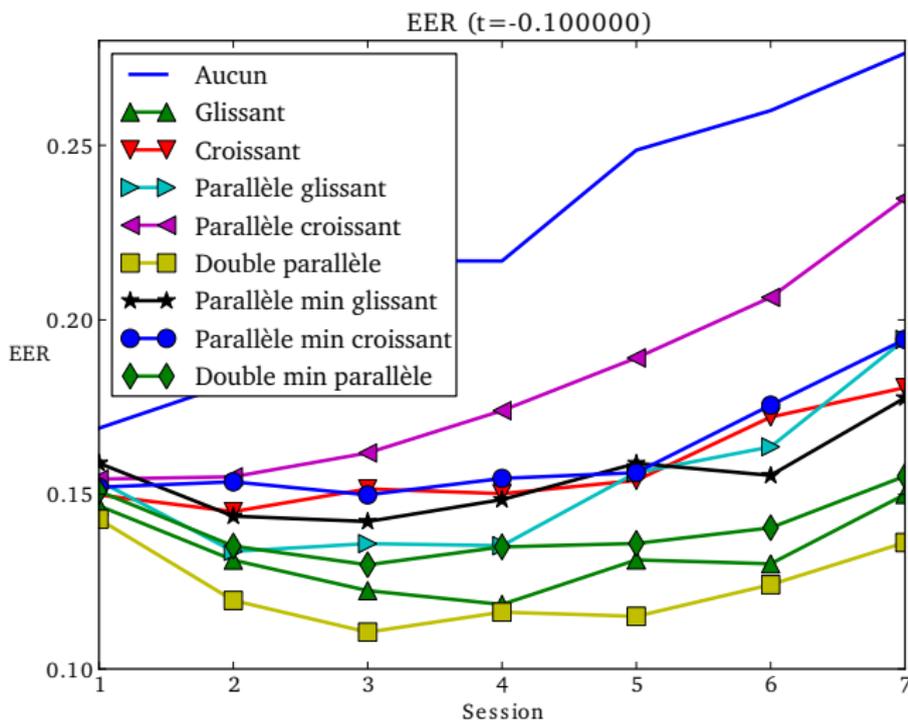
Protocole - Configuration du système

Modalité	Dynamique de frappe au clavier
Méthode d'authentification	Calcul de distance <small>(de Magalhaes et al., 2005)</small>
Décision de mise à jour	Double seuillage en ligne et semi-supervisée
Seuil de mise à jour	Fixé empiriquement
Systèmes de mise à jour	Aucun, fenêtre glissante, fenêtre croissante
Configurations	<ul style="list-style-type: none">● (Aucun, Glissante)● (Aucun, Croissante),● (Glissante, Croissante)
Fusion	Valeur moyenne, valeur minimum

Protocole - Configuration de l'évaluation

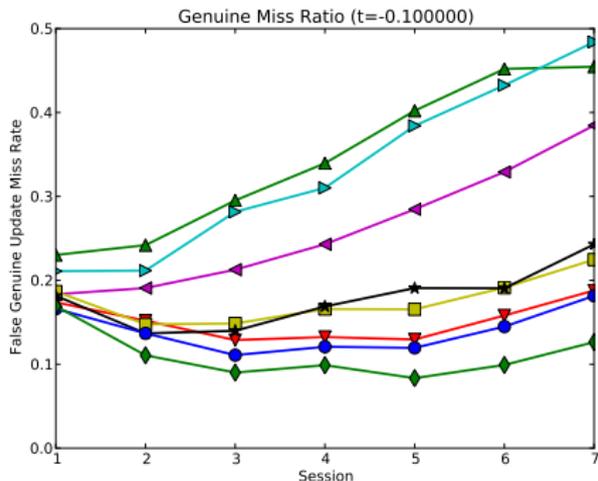
Jeux de données	DSL2009, GREYC2009
Nombre de sessions	8 pour DSL2009, 5 pour GREYC2009
Respect de la chronologie	Oui
Ordre de présentation	Aléatoire
Quantité d'imposteurs	30% d'exemples d'impostures (100 exécutions moyennées)
Calcul de l'évaluation	En ligne
Métriques d'évaluation	EER, FNMR, FMR, IUSR, GUMR (scores de la session actuelle, pas de moyenne avec les sessions précédentes)

Performances de reconnaissance

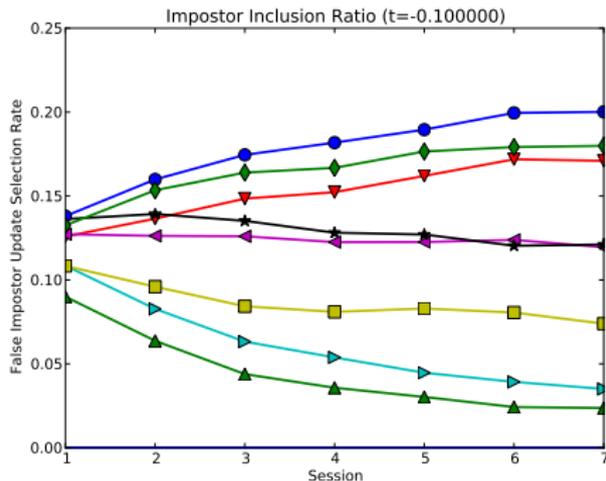


Évolution de l'EER au cours des sessions

Performances de mise à jour



Oublis de clients



Inclusion d'imposteurs

- En général, **compromis** entre :
 - inclure des imposteurs et oublier des clients
- Sauf pour **Double parallèle**

Discussion

L'évolution parallèle permet de

- **Lisser** les **erreurs** de mise à jour
- Obtenir une **meilleure résistance** au vieillissement

Cependant

- Méthode généralisable ?
- Bien que non testé, le système peut être utilisé dans d'autres contextes :
 - hors ligne, supervisé, classifieur online...

Discussion

L'évolution parallèle permet de

- **Lisser** les **erreurs** de mise à jour
- Obtenir une **meilleure résistance** au vieillissement

Cependant

- Méthode généralisable ?
- Bien que non testé, le système peut être utilisé dans d'autres contextes :
 - hors ligne, supervisé, classifieur online. . .

5 Conclusion

- Rappel et discussion du travail effectué
- Valorisation des travaux

Dynamique De Frappe

Travail réalisé

- Création de 2 bases de données conséquentes
- 1 méthode d'authentification par mot de passe partagé
- Caractérisation des bases existantes
- Analyse des facteurs qui influencent les performances
 - Mode d'acquisition
 - Propriétés des mots de passe

Perspectives

- Mise au point de **règles** de sélection de mot de passe maximisant les performances
- Éviter le **re-jeu** (Thèse de Syed Idrus Syed Zulkarnain)

Multibiométrie

Travail effectué

- Approximation de l'EER
- Définition de nouvelles fonctions de fusion
- Utilisation de la programmation génétique

Perspectives

- Faire de la **sélection** de **caractéristiques**
- Diminuer la **complexité** d'approches basées sur la **programmation génétique**

Biométrie douce

Travail effectué

- Texte fixe :
 - Reconnaissance du genre et augmentation du taux d'authentification en utilisant la biométrie douce
- Texte libre :
 - Détection du genre et de catégorie d'âge

Perspectives

- Compléter cette étude
 - Nécessité de créer des **bases** plus conséquentes
 - Recherche de **catégories** plus utiles et **performantes**
 - Utilisation dans un **sac de biométrie douce** (Dantcheva et al., 2011)

Mise à jour de la référence biométrique

Travail réalisé

- Analyse des propriétés des systèmes de mise à jour
- Analyse de l'impact de la méthode d'évaluation
- Mise au point d'un système hybride meilleur que l'auto-apprentissage

Perspectives

- Utilisation des données d'**imposteurs**
- Prédiction de la **dévi**ation de la donnée biométrique
- Il est nécessaire d'établir des **protocoles** communs d'**évaluation** de systèmes de mise à jour de la référence

Valorisation des travaux

Type de publication	Quantité
Chapitre de livre	4
Article de revue avec comité de rédaction	5
Conférence internationale avec comité de sélection	14
Conférence internationale sans comité de sélection	1
Conférence nationale avec comité de sélection	4
Rapport technique	4
Logiciel	2

Merci pour votre attention

Extrait de l'état de l'art

Étude	A	B	C	D	E	FAR	FRR
Bleha et al., 1990	8 semaines	36	30	yes	yes	2.8%	8.1%
Obaidat at al, 1997	8 semaines	15	112	no	no	0%	0%
Monrose et al., 1997	7 semaines	42	??	no	no	??	20%
Rodrigues et al., 2006	4 sessions	20	30	??	no	3.6%	3.6%
Revet et al., 2006	4 semaines	8	12	??	??	5.58%	5.58%
Hocquet et al., 2007	??	38	??	??	no	1.7%	2.1%
Revet et al., 2007	14 jours	30	10	??	no	0.15%	0.2%
Hosseinzadeh et al., 2008	??	41	30	no	no	4.3%	4.8%
Killourhy et al., 2009	8 sessions	51	200	yes	no	9.6%	9.6%

Données utilisées

Données brutes possibles

- Clavier normal : heure de pression
- Clavier spécifique : force exercée *(Eltahir et al., 2008)*
- Caméra : positionnement des doigts *(Nishiuchi et al., 2010)*
- Microphone : son produit par les touches *(Dozono et al et al., 2007)*
- Ordinateur portable récent : vibrations capturées par le disque dur *(Lopatka et al., 2009)*

Performance mobile - Projet M2 Recherche

		Test 1	Test 2	Test 3	Test 4
Données					
	Pression	•		•	•
	Temps de vol		•	•	•
	Temps d'appui		•	•	•
	Coordonnées d'appui				•
	Décalage du doigt				•
Mot de passe	EER				
23648	Statistique	36.9 ⁰ %	29.9 ⁰ %	29.3 ⁰ %	28.2 ⁰ %
	Revet	37.2 ⁰ %	29.9 ⁰ %	28.1 ⁰ %	27.8 ⁰ %
	GMM	35.9 ⁰ %	29.5 ⁰ %	29.4 ⁰ %	29.3 ⁰ %
16573049	Statistique	36.1 ⁰ %	27.7 ⁰ %	27 ⁰ %	27 ⁰ %
	Revet	35.7 ⁰ %	26.7 ⁰ %	26.6 ⁰ %	25.7⁰%
	GMM	33.2 ⁰ %	29.9 ⁰ %	28.8 ⁰ %	30.6 ⁰ %

(Ulric, 2011)

Illustration EER approximé

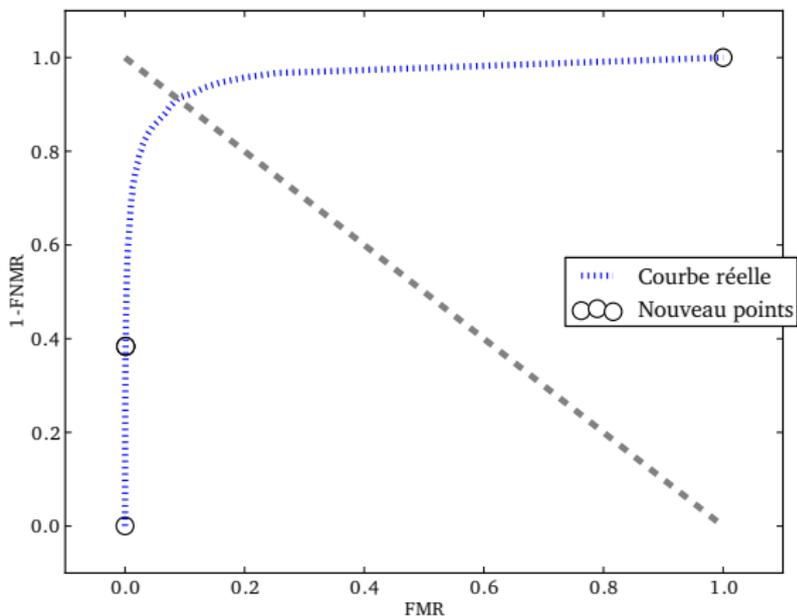


Illustration EER approximé

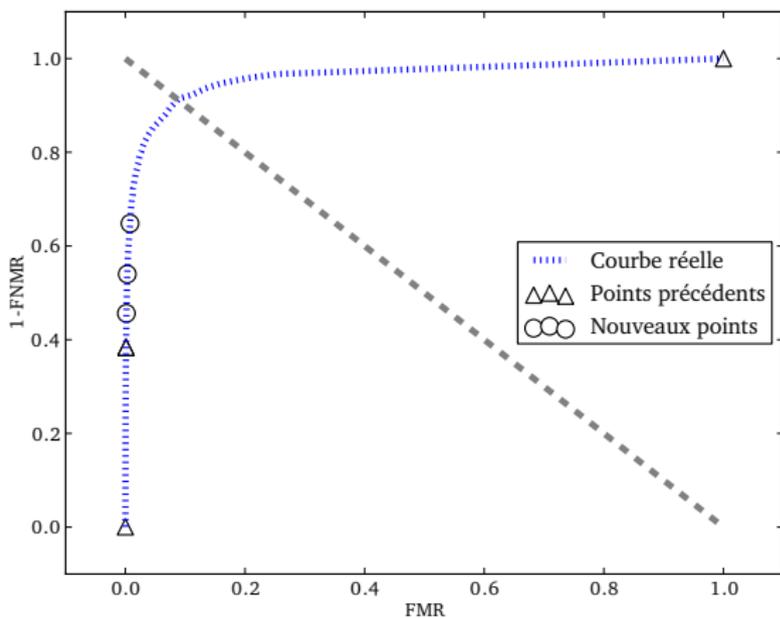


Illustration EER approximé

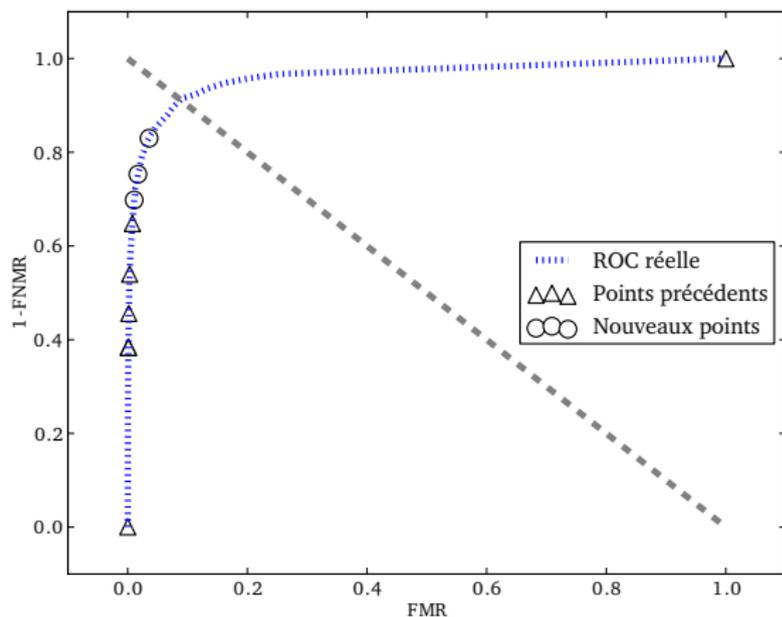


Illustration EER approximé

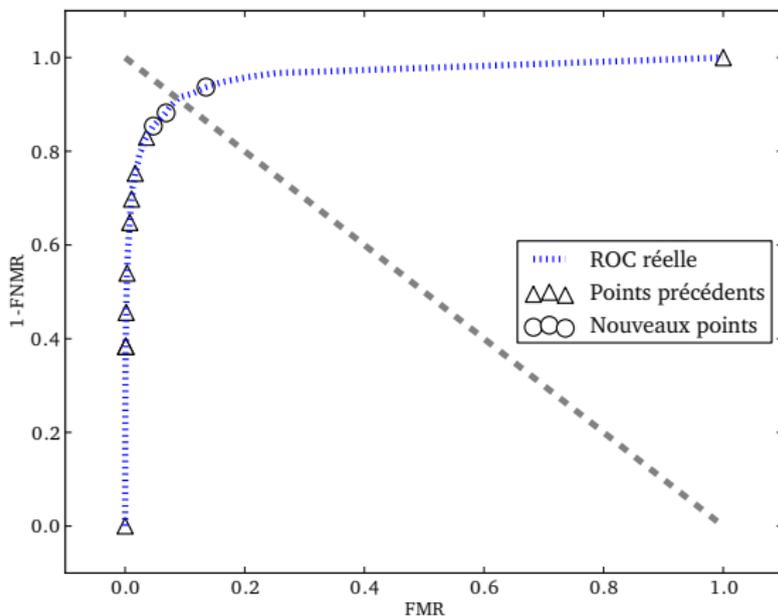


Illustration EER approximé

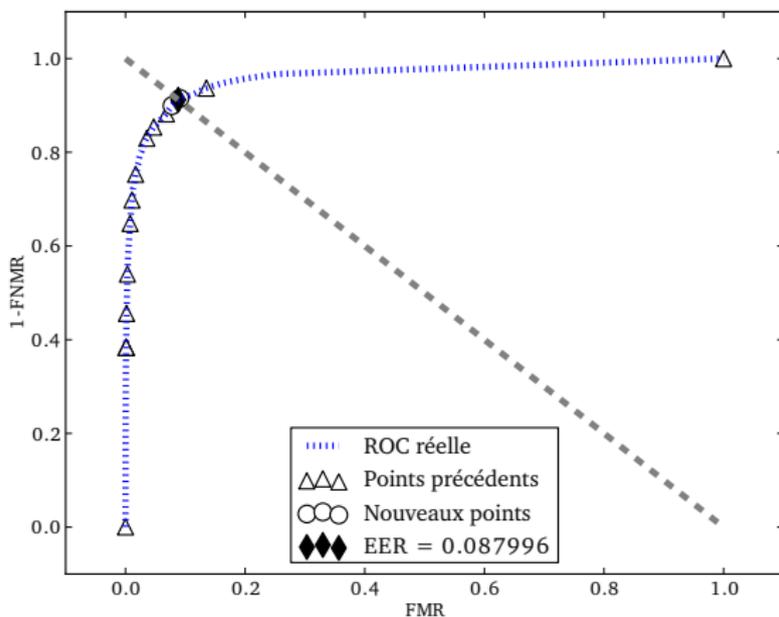
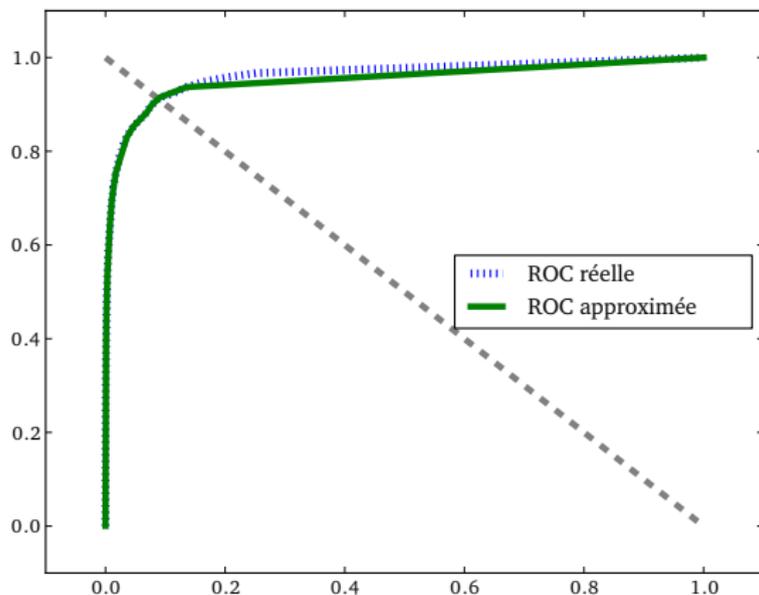


Illustration EER approximé



EER approximé/configurations

LABEL	ERROR (%)	TIME (ms.)	COMP.
polyto_3_0.010	0.30	110	11
polyto_3_0.005	0.07	139	14
polyto_3_0.003	0.07	140	14
polyto_4_0.010	0.40	140	15
polyto_4_0.005	0.20	149	16
polyto_4_0.003	0.10	169	18
polyto_5_0.010	0.30	150	16
polyto_5_0.005	0.07	190	20
polyto_5_0.003	0.07	179	20
polyto_6_0.010	0.40	140	15
polyto_6_0.005	0.10	179	19
polyto_6_0.003	0.10	179	19
polyto_7_0.010	0.07	190	21
polyto_7_0.005	0.07	190	21
polyto_7_0.003	0.07	200	21

Caractéristiques extraites de la base Bello et al.

_ représente le caractère espace

- « e_ »
- « a_ »
- « _d »
- « de »
- « la »
- « _l »

Performance

Méthode (EER)	Sans	Manuel	Automatique	Semi-automatique
label/template	10,65%	<u>9,74%</u>	10,05%	9,99%
label/score	<u>10,65%</u>	10,65%	10,74%	27,64%
score/template	10,65%	<u>9,74%</u>	10,48%	10,58%
score/score	10,65%	<u>7,64%</u>	8,45%	17,52%
Meilleure	10,65%	<u>7,64%</u>	8,45%	9,99%
Gain	x	<u>28,26%</u>	20,67%	6,20%

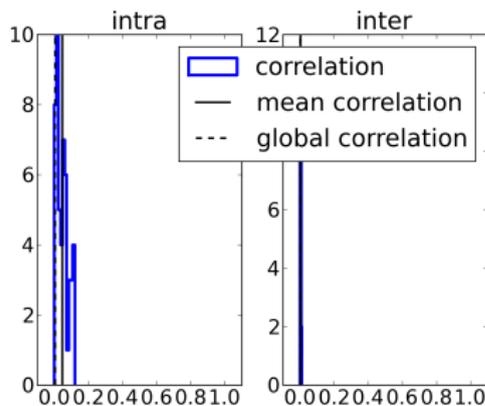
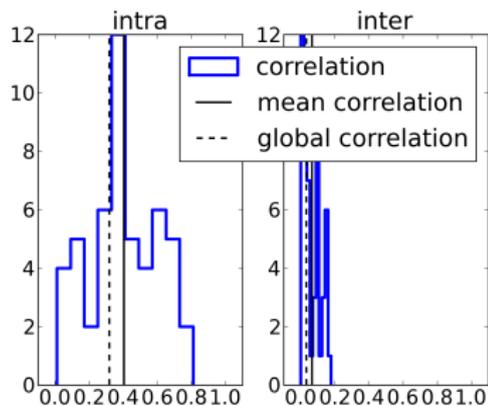
Auto-apprentissage (Rolli et al., 2006)

- 1 La référence est générée en entraînant le classifieur avec les données étiquetées (DL)
- 2 Un ensemble de données non étiquetées (DU) est collecté tout au long de l'utilisation du système biométrique
- 3 Le classifieur est utilisé pour étiqueter les données incomplètes (DU)
- 4 Les données étiquetées avec un fort degré de confiance sont ajoutées aux données d'apprentissage (DL)
- 5 Le classifieur est ré-entraîné sur l'ensemble des données étiquetées (DL) et la procédure est répétée à l'étape 3 jusqu'à satisfaire un critère d'arrêt

Co-apprentissage (Rattani et al., 2008)

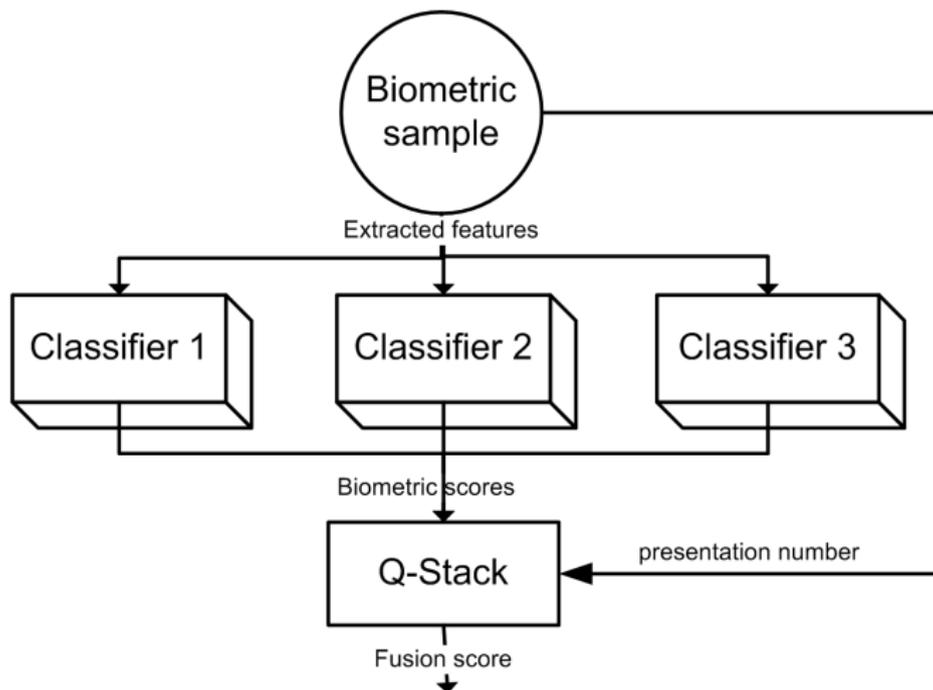
- 1 La référence de chaque modalité est calculée à l'aide de l'ensemble DL .
- 2 Un ensemble de données non étiquetées (DU) est collecté tout au long de l'utilisation du système biométrique
- 3 Chacun des deux classifieurs est utilisé pour étiqueter les données incomplètes (DU)
- 4 Les données étiquetées avec un fort degré de confiance (par un des classifieurs) sont ajoutées aux données d'apprentissage (DL)
- 5 Les classifieurs sont ré-entraînés sur l'ensemble des données étiquetées (DL) et la procédure est répétée à l'étape 3

Corrélation temps/performance (IHMSP2012)



(Giot et al., 2012)

Qstack - Schéma



Qstack - Comportement

	GREYC [?]			DSL2009 [?]		
	Normal	Stacked	Gain (%)	Normal	Stacked	Gain (%)
SVM/gr. 1	[14.12-14.85]	[7.14-21.42]	1.43	[27.47-28.19]	[15.75-47.26]	-13.22
SVM/gr. 2	[13.03-13.98]	[6.68-20.04]	1.06	[15.45-16.05]	[7.77-23.32]	1.28
SVM/gr. 3	[14.83-16.11]	[8.23-24.68]	-6.36	[38.31-38.70]	[20.35-61.05]	-5.69

Séparation linéaire

$$y(\mathbf{x}) = \mathbf{w}^T \cdot \phi(\mathbf{x}) + b \quad (3)$$

Apprentissage

$$\forall n, t_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b) \geq 1 \quad (4)$$

Distance entre marge et point

$$d(\mathbf{x}) = \frac{|y(\mathbf{x})|}{\|\mathbf{w}\|} \quad (5)$$

Maximisation marge = minimisation de

$$\operatorname{argmin}_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (6)$$

Marge souple

$$\forall n, t_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b) \geq 1 - \xi_n \quad (7)$$

Noyau Gaussien

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \text{ avec } \gamma > 0 \quad (8)$$