Vérification du scripteur

Ameur Bensefia, Thierry Paquet, Laurent Heutte

▶ **To cite this version:**

Ameur Bensefia, Thierry Paquet, Laurent Heutte. Vérification du scripteur. Jun 2004. sic_00001191

HAL Id: sic_00001191
https://archivesic.ccsd.cnrs.fr/sic_00001191

Submitted on 7 Dec 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vérification du scripteur

Ameur BENSEFIA – Thierry PAQUET – Laurent HEUTTE

Laboratoire PSI, CNRS FRE 2645 – Université de Rouen
Place E. Blondel, UFR des Sciences et Techniques
Université de Rouen, F-76 821 Mont-Saint-Aignan cedex

{Ameur.Bensefia,Thierry.Paquet,Laurent.Heutte}@univ-rouen.fr

Résumé *: Dans cette communication nous nous intéressons à la tâche de vérification du scripteur. Il s'agit de comparer deux échantillons d'écritures inconnues et de décider s'ils ont été produits par le même scripteur ou non. Les écritures sont initialement caractérisées par les graphèmes produits par un algorithme de segmentation. Les échantillons d'écriture sont ensuite analysés de manière différente selon qu'ils représentent des textes de quelques lignes d'écriture ou au contraire un seul mot. Les textes sont projetés dans un espace de caractéristiques commun aux deux scripteurs. La statistique du critère d'information mutuelle entre l'ensemble des 2 scripteurs et l'ensemble des caractéristiques permet de construire un test d'hypothèse fiable. Dans le cas d'échantillons de petite taille (des mots) on utilise la distance de Levenstein pour construire un second test d'hypothèse. Les deux approches sont évaluées sur des bases du PSI et sur la base IAM.*

Mots-clés *:* Vérification du scripteur, information mutuelle, distance de Levenstein.

1 Introduction

Les systèmes biométriques conçus pour la vérification d'individus se fondent sur une hypothèse selon laquelle la variation de caractéristiques appartenant à un même individu (intra) est moins significative que la variation de caractéristiques entre deux individus différents (inter). Un certains nombre d'identifiants biométriques sont réputés pour disposer de cette propriété (empreintes digitales, rétiniennes, forme du visage…). Dans ce domaine on distingue les caractéristiques physiques telles que celles que nous venons de citer et les caractéristiques comportementales telles que la signature ou la parole par exemple [Jai 99]. Dans cette communication nous évaluons l'aptitude de l'écriture manuscrite à constituer un identifiant biométrique comportemental. Il ne s'agit pas ici de considérer l'écriture comme un nouvel identifiant biométrique qui pourrait se substituer aux identifiant physiques connus pour leur robustesse, mais plutôt de montrer dans quelle mesure l'écriture peut constituer une caractéristique de substitution lorsqu'elle représente la seule marque de l'individu dont on dispose. Par exemple sur un manuscrit ancien ou de manière générale une correspondance.

Dans ce cadre, la tâche de vérification du scripteur consiste à authentifier le scripteur d'un document. La plupart du temps cette tâche est réalisée par un expert et est empreinte d'une subjectivité importante. En tout état de cause, la confiance que l'on peut associer à une décision de ce type n'est pas scientifiquement quantifiable. Des travaux récents ont proposé une méthodologie d'analyse scientifique des écritures notamment pour la tâche de vérification du scripteur [Cha 00]. Il faut noter que cette tâche a été rarement étudiée au regard des nombreuses études qui ont concerné la tâche d'identification. Ceci est sans doute dû au fait que la vérification implique des processus de décisions locales qui dépendent généralement du contenu textuel. En effet, on se ramène généralement à devoir comparer les formes possibles d'un caractère ou d'un mot spécifique présent sur les documents étudiés. C'est pourquoi l'automatisation complète de cette tâche semble peu réaliste car elle solliciterait un système de reconnaissance automatique. De ce fait, la vérification du scripteur est généralement menée en fonction d'un contenu textuel.

La première étape dans la mise en œuvre d'un système de vérification du scripteur nécessite de définir une mesure de similitude entre deux écritures manuscrites. La seconde étape est une décision entre deux classes : intra et inter. A cet effet un test d'hypothèse se prête parfaitement à l'application, car il permet de quantifier l'erreur de fausse acceptation et de faux rejet.

Dans cette communication, nous proposons d'aborder la tâche de vérification du scripteur de 2 façons différentes - indépendamment du contenu textuel pour des échantillons de grande taille (blocs de textes) - en fonction du contenu textuel pour de petits échantillons (mots manuscrits). La première partie est consacrée à la vérification du scripteur pour des documents manuscrits. L'approche repose sur un critère d'information mutuelle et sur un test d'hypothèse pour la prise de décision. Deux bases ont été utilisées pour évaluer la méthode : la *Base_PSI* composée de 88 scripteurs, cette base a servi de base d'apprentissage, et la *Base_IAM* composée de 150 scripteurs qui a servie de base de test.

Dans la seconde partie de cete communication nous abordons la vérification du scripteur sur des mots manuscrits. Pour cela nous mettons en œuvre un calcul de distance d'édition entre chaînes manuscrites inspiré

de l'algorithme de Wagner et Fisher [Wag 74]. Nous avons construit un test d'hypothèses pour élaborer la décision. La méthode est évaluée sur une base de 100 mots écrits par 20 scripteurs différents et construite au PSI.

2 Vérification du scripteur sur des documents manuscrits

Rappelons que la tâche de vérification du scripteur qui nous intéresse dans ce cadre consiste à décider si deux écritures présentes sur 2 documents différents peuvent être du même scripteur ou non. Pour cela il convient de choisir un ensemble de caractéristiques graphiques pertinent capable de décrire n'importe quelle écriture. Il est a priori difficile voire impossible de prétendre à une telle exhaustivité. Pour contourner cette difficulté, la méthode repose sur un ensemble de caractéristiques adaptatif aux écritures examinées dont nous rappelons brièvement le principe de détection. Pour plus de précisions nous renvoyons à la lecture de [NOS 02]. Le document est segmenté en lignes d'écritures et chaque ligne d'écriture est segmentée en graphèmes à partir de l'analyse des contours supérieurs des composantes connexes. Cette étape produit un ensemble de formes élémentaires qui sont soit des caractères, soit des parties de caractères, soit enfin des caractères collés. De ce fait, cet ensemble peut ensuite donner lieu à une description en bi-grammes ou tri-grammes de graphèmes si l'on souhaite reconstruire exactement les caractères manuscrits du texte.

Dans le cadre de la vérification des scripteurs qui nous préoccupe ici, les textes soumis à l'analyse sont segmentés, et l'ensemble des graphèmes obtenus sur les deux documents est soumis à une phase de classification non supervisée. Cette étape permet de définir des groupes de formes plus ou moins fréquents sur les deux documents qui vont constituer les caractéristiques de l'analyse. De ce fait la méthode n'est pas directement dépendante d'un ensemble de caractéristiques pré-établi et figé et c'est précisément ce qui la rend adaptable à des écritures quelconques inconnues du système. Nous décrivons maintenant les 2 étapes essentielles qui conduisent à la décision.

2.1 Similarité entre scripteurs

Intuitivement il semble naturel de considérer que deux écritures sont d'autant plus proches qu'elles se décomposent sur les mêmes caractéristiques et ceci avec la même régularité (fréquence). Ainsi on peut considérer qu'en moyenne 2 écritures sont identiques si les distributions statistiques des caractéristiques qui les composent sont voisines. Nous avons construit un critère d'information mutuelle pour quantifier la relation qui existe entre d'une part un ensemble de caractéristiques discrètes (les groupes de graphèmes) et l'ensemble des deux scripteurs soumis à l'analyse.

Plus formellement, si S et G désignent respectivement l'ensemble des 2 scripteurs (S_1 et S_2) et G l'ensemble des groupes de graphèmes g_i on note :

$S = \{ S_1, S_2 \}$ $\quad$ $G = \{ g_1, g_2, g_3, \ldots . g_N \}$

L'information mutuelle mesure la dépendance entre l'ensemble des deux scripteurs *S* et l'ensemble des caractéristiques *G*. Elle est définie par la relation :

$$I_M(G,S)=H(G)-H(G/S)$$

où $H(G)$ désigne l'entropie de Shannon définie par :

$$H(G)=-\sum_{i=1}^{card(G)} P(g_i)\, H(G=g_i) = -\sum_{i=1}^{card(G)} P(g_i)\, log_2[P(G=g_i)]$$

et $H(G/S)$ désigne l'entropie conditionnelle définie par:

$$H(G/S) = -\sum_{i=1}^{card(G)} \sum_{j=1}^{card(S)} P(S_j)\, P(g_i/S_j)\, log_2[P(g_i/S_j)]$$

De faibles valeurs de l'information mutuelle indiquent une indépendance entre les deux variables aléatoires S et G. L'indépendance entre *S* et *G* reflète le fait que l'ensemble des caractéristiques *G* est distribué de la même façon sur les deux documents et dans notre cas est un bon indicateur pour considérer les 2 scripteurs identiques. Dans le cas contraire, il y a une dépendance entre *S* et *G* et ceci révèle que les deux scripteurs n'utilisent pas les caractéristiques (les graphèmes) de la même façon.

2.2 Expérimentations

Pour attester de la pertinence de ce critère nous avons mené deux séries de tests distinctes sur deux bases d'écritures : une base des 88 scripteurs (*Base_PSI*) et une base de 150 scripteurs (*Base_IAM*) (http://www.iam.unibe.ch/~fki/iamDB) [ZIM 02]. Ces deux bases sont constituées de textes écrits par des scripteurs différents. La base PSI correspond à la recopie d'un même texte tandis que la base IAM comprend des textes différents pour chaque scripteur (productions libres). Dans les deux cas les textes ont été coupés en deux parties, de façon à disposer de deux échantillons d'écriture de chaque scripteur. La figure 1 représente la distribution du critère d'information mutuelle sur la *Base_PSI*. Nous constatons une distribution de ce critère nettement différente dans les cas intra (a et c) et inter-scripteurs (b et d) et ceci pour les graphèmes (a,b) et les bigrammes (c,d). La distribution des tri-grammes, que nous n'avons pas représenté par manque de place, est nettement moins intéressante notamment parce qu'il y a peu de répétitions significatives de ces formes.

2.2.1 Décision

La vérification du scripteur est un problème de prise de décision entre deux classes : intra et inter-scripteur. L'élaboration de la meilleure décision nous conduit à modéliser la distribution statistique du critère de décision pour les deux classes considérées. Nous avons choisi d'utiliser la *Base_PSI* comme base d'apprentissage. Le système a donc ensuite été évalué sur la plus grande base : la *Base_IAM*.

La modélisation du problème d'identification du scripteur correspond à la modélisation de deux hypothèses :

– H_0 : Les deux écritures sont l'œuvre d'un même scripteur

– H_1 : Les deux écritures sont l'œuvre de scripteurs différents

Il convient alors de modéliser le risque associé à chaque décision. Dans un test d'hypothèses, la décision aboutira à choisir entre H_0 ou H_1. Nous pouvons, par conséquent, commettre deux erreurs : l'erreur de première espèce notée α ou erreur de fausse acceptation et l'erreur de seconde espèce notée β ou erreur de faux rejet [Sap 90]. La puissance du test 1-β indique la capacité du test à identifier les cas où l'hypothèse nulle H_0 est fausse. Elle diminue avec l'erreur de première espèce α . Pour une valeur de α donnée, la puissance du test augmente avec la taille de l'échantillon.

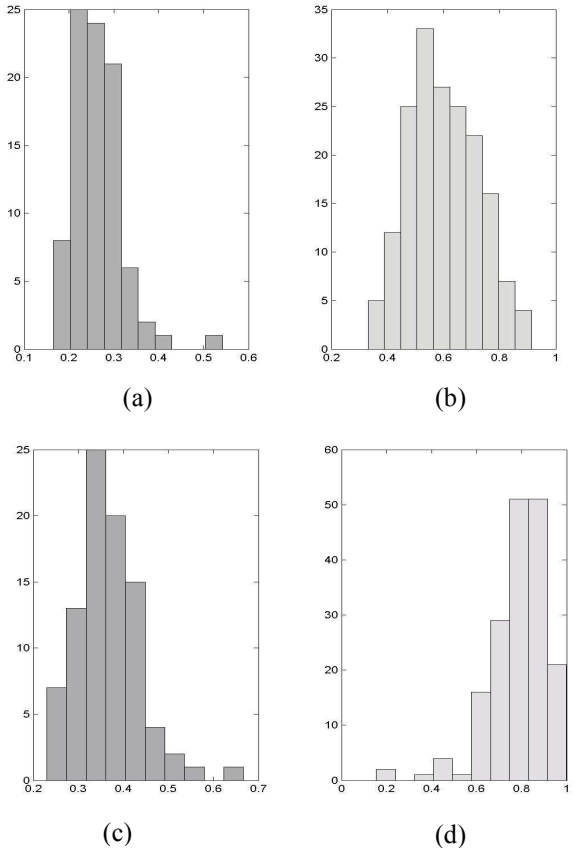

FIG. 1 - Distribution du critère d'information sur la base PSI, sur les graphème a,b), et sur les bi-grammes c,d). en intra a,c) et inter-scripteurs b,d).

2.2.2 Choix de la statistique

Plusieurs tests de conceptions très différentes sont disponibles pour soumettre à une épreuve de vérité une hypothèse principale. Dans de tels cas, le test qui fournit l'erreur β la plus petite, pour une même valeur de α, est par définition le plus puissant (celui ayant la plus grande valeur de la puissance de test 1- β). En effet, il peut détecter les plus petites différences entre les populations sans pour autant augmenter l'erreur de première espèce. Décider du choix du test statistique le plus approprié aux hypothèses est un aspect central des statistiques. Chaque statistique de test possède une distribution de probabilité. Par exemple, pour un échantillon de grande taille et de variance connue, on peut opter pour une statistique fondée sur une distribution normale de moyenne μ et d'écart type σ. Si l'écart type de la population est inconnu, ce qui est le plus souvent le cas, alors la statistique de test la plus appropriée est fondée sur une distribution de *t Student*. La loi de *t Student* est souvent utilisée quand le nombre d'individus qui composent la population est inférieur à 30, car au-delà la loi de *t Student* converge vers une loi normale. De ce fait nous avons évalué notre système de vérification du scripteur selon une statistique de loi normale. Il se peut que d'autres distributions soient plus adaptées mais l'ajustement avec une loi normale a semblé suffisant (figure 2).

2.2.3 Résultats

Le tableau 1 dresse un récapitulatif des différentes valeurs de α, $1-\beta$, et du seuil de décision *Vc* qui sera associé à ces 2 erreurs sur la *Base_PSI*. Notons que du fait que nous disposons de la distribution de l'hypothèse H_1, nous pouvons nous défaire du consensus historique du choix de la valeur de α (2.5% ou 5%). En effet, d'après la théorie des lois Bayésiennes [Dud 00] la valeur optimale de cette erreur, qui permet de minimiser les deux régions d'erreurs, se trouve à l'intersection des deux distributions.

Comme nous l'avons indiqué précédemment, nous avons gardé la *Base_IAM* pour l'évaluation de notre système de vérification d'écritures. La première série de test consiste à mesurer les similitudes intra scripteur, puis les similitudes inter-scripteurs afin d'appliquer notre règle de décision. Les résultats obtenus sont résumés dans la tableau 2 ci-dessous.

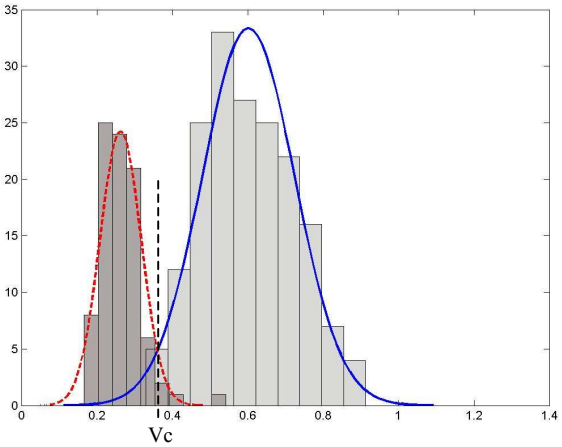

FIG. 2 - Ajustement des distributions en intra et inter-scripteurs avec une loi normale sur la base PSI.

BASE_PSI	α	1–β	*Vc*
GRAPHEMES	3.5%	97.5%	0.36164
BI-GRAMMES	3.4%	98.5%	0.4984

TAB. 1 - Récapitulatif des différentes valeurs de α, 1-β et Vc avec une statistique de loi normale calculées sur la Base_PSI.

BASE_IAM	ERREUR I (α)	ERREUR II (1–β)
GRAPHEMES	4%	96%
BI-GRAMMES	10.66%	96.66%

TAB. 2 - Evaluation de l'erreur de 1er et 2nd espèce sur la Base_IAM.

2.2.4 Discussion

Les résultats obtenus semblent particulièrement prometteurs à plusieurs égards. Tout d'abord le choix d'une représentation locale fondée sur les graphèmes segmentés semble tout à fait pertinent puisqu'il permet un niveau de description proche des caractères sans toutefois nécessiter une étape de reconnaissance ou une étape de segmentation manuelle comme le propose [SRI 01]. D'autre part, il est tout à fait intéressant et fort encourageant d'obtenir un niveau de performance sur la *Base_IAM* du même ordre que sur la *Base_PSI* sur laquelle a été effectué l'apprentissage du test d'hypothèse. On remarque cependant la valeur plus importante que prévue de l'erreur de première espèce sur les bi-grammes. Ceci peut s'expliquer par la faible représentation des groupes de bi-grammes sur la *Base_IAM* parce que sur cette base les textes sont en général plus courts que sur la *Base_PSI*. On dispose donc de beaucoup moins d'information pour identifier le scripteur dans ce cas.

3 Vérification à partir de mots manuscrits

S'attaquer à la vérification du scripteur par l'examen d'échantillons de mots manuscrits impose de s'attacher au contenu textuel afin de combler le manque d'information. Rappelons, que les experts en écriture, lors d'un processus de vérification de documents, s'attachent à comparer les caractéristiques extraites sur les mêmes caractères ou sur les mêmes mots des deux documents [MOR 00][JAR 99]. En travaillant sur des documents manuscrits nous avons pu nous afranchir de cette restriction et proposer l'automatisation d'un test de vérification d'écriture. L'approche que nous proposons maintenant pour l'identification à partir de mots manuscrits s'affranchit de la connaissance des caractères dans les mots. En revanche, nous nous plaçons maintenant dans la situation de comparer des mots identiques. On peut alors rapprocher la comparaison de mots manuscrits aux travaux consacrés à la vérification de signatures [BOV 03], ou à certaines approches de reconnaissance de l'écriture cursive [PAR 96]. Nous nous en sommes inspirés pour proposer un nouveau critère basé sur le calcul d'une distance entre deux mots manuscrits.

3.1 Distances entre chaînes manuscrites

Notre problématique de vérification de mots manuscrits peut être considéré comme un problème de calcul de distance d'édition entre 2 chaînes manuscrites dès lors qu'on est capable de segmenter les écritures en graphèmes au moyen d'un quelconque procédé de segmentation. C'est précisément notre cas, comme nous avons eu l'occasion de le préciser dans la partie 2. On formule alors le problème de la manière suivante : considérons deux mots manuscrits *X* et *Y* segmentés

$$X = x_0\, x_1 \,..\, x_n \qquad\qquad Y = y_0\, y_1 \,..\, y_m$$

où les x_i, y_j représentent respectivement, les graphèmes de la chaîne *X* et les graphèmes de la chaîne *Y*. Nous définissons le coût de transformation d'un graphème x_i en un graphème y_j par une mesure de similarité entre graphèmes (dans notre cas la corrélation). De fait l'algorithme de calcul de distance entre mots manuscrit s'écrit simplement :

```
D(0, 0) := sim(x0, y0)
Pour i :=1 à n  Faire
       D(i, 0) := D(i-1, 0) + sim(xi, y0)
Pour j =1 à m  Faire
       D(0, j) := D(0, j-1) + sim(x0, yj)

Pour  i :=1 à n  Faire
       Pour j := 1 à m Faire
              M1 := D(i-1, j-1) +  sim(xi, yj)
              M1 := D(i-1, j)  +  sim(xi, yj)
              M1 := D(i, j-1)  +  sim(xi, yj)
              D(i, j) := Max(M1, M2, M3)

Distance(X, Y) = D(n, m)/ (n+m)
```

où $sim(x_i, y_j)$ est la mesure de corrélation entre le graphème x_i et le graphème y_j

En faisant ce choix, les valeurs maximales du coût de transformation de X en Y traduisent une faible variabilité entre les deux chaînes, et nous conduit à privilégier le fait que les deux mots considérés sont issus de la même main. Inversement, de faibles valeurs du coût final de la transformation seront significatifs d'une différence importante entre les deux scripteurs des deux mots analysés.

3.2 Expérimentations

Les expérimentations menées dans le cadre de la vérification de mots manuscrits ont nécessité la conception d'une nouvelle base. Nous avons demandé à 20 scripteurs différents de recopier 5 occurrences du mot « manuscrit ». Quelques échantillons de cette base sont présentés sur la figure 3. Nous avons utilisé 60 distances intra-scripteur et 60 distance inter-scripteur pour évaluer

les distributions intra et inter-scripteurs que nous avons également modélisées avec des lois normales.

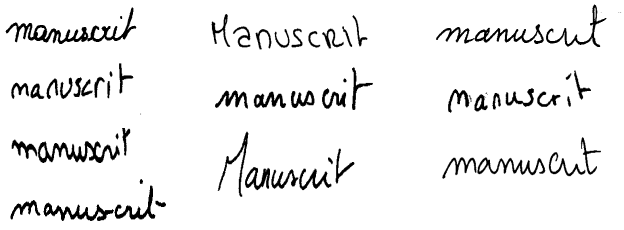

FIG. 3 - Quelques échantillons utilisés pour la vérification de scripteurs.

Nous avons respecté le même procédé que celui adopté pour la vérification des scripteurs sur des documents et formalisé le problème selon deux hypothèses H_0 et H_1 et déterminé les deux erreurs qui leur sont associées. Les ajustements réalisés sur les distributions permettent de déterminer les régions de rejet et d'acceptation selon un seuil de décision V_c (figure 4). Nous constatons une grande zone de recouvrement entre les deux distributions, ce qui conduit naturellement à un risque d'erreur plus important que dans le cas précédent. On estime en effet la puissance du test à 78%, pour une erreur de première espèce α de 15.1%. Ces résultats semblent tout à fait naturels du fait de la faible quantité d'information disponible pour construire ce second test.

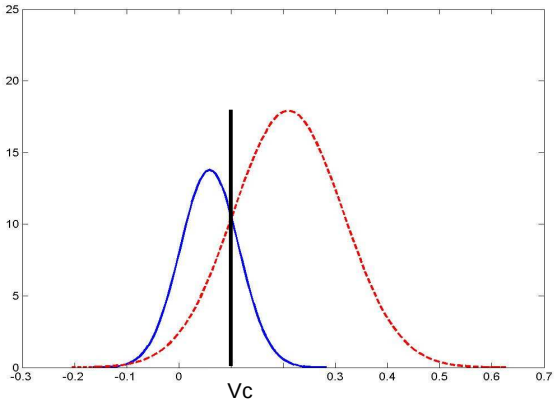

FIG. 4 - Distribution des distances inter et intra-scripteurs pour la vérification de mots manuscrits.

Pour évaluer la pertinence de ce nouveau système dédié à la vérification des mots, nous lui avons soumis 60 paires de mots provenant du même scripteur et 60 autres paires de mots provenant de scripteurs différents. Les résultats sont présentés dans le tableau 3. L'erreur de première espèce est du même ordre que celle obtenue sur la base PSI. L'erreur de seconde espèce en revanche a diminué fortement (10% au lieu de 21%). Ceci est principalement dû à la faible taille des échantillons utilisés qui ne permet pas au scripteur de faire ressortir immédiatement ses traits et ses caractéristiques personnelles.

BASE_MOTS	ERREUR I	ERREUR II
ERREURS SUR LA BASE PSI	15.1%	79%
ERREUR OBSERVEE SUR LA BASE IAM	16.66%	90%

TAB. 3 - Performances de vérification sur la base de mots avec une statistique de loi normale

4 Conclusion

Dans cette communication nous nous sommes intéressés à la problématique de la vérification du scripteur. Elle constitue une approche complémentaire et nécessaire pour toute approche d'identification, car elle offre l'unique moyen de rejeter un individu (scripteur) lorsque ce dernier est inconnu du système. Nous nous sommes intéressés à la vérification de documents manuscrits, indépendamment de leur contenu textuel, et à la vérification de mots manuscrits où nous avons tenté de nous affranchir de la connaissance des caractères. Dans les deux cas, nous avons construit un critère discriminant et une règle de décision basée sur un test d'hypothèse. Ceci permet de quantifier les erreurs de décisions (erreur de fausse acceptation et erreur de faux rejet).

Les résultats de vérification obtenus à partir de l'analyse de textes sont nettement meilleurs que ceux obtenus à partir de l'analyse de mots ce qui semble tout à fait naturel. Nous avons montré qu'il est possible de vérifier les écritures de deux documents sans avoir besoin de connaître préalablement les propriété de ces écritures. Les tests ont été réalisés sur une base de 150 scripteurs inconnus du système et ont permit d'obtenir une erreur de fausse acceptation de l'ordre de 4%. Les résultats obtenus en vérification sur des mots manuscrits sont de l'ordre de 15% de fausse acceptation ce qui est tout à fait honorable sans utiliser la connaissance des caractères dans les mots.

Ces résultats mettent en évidence la capacité à caractériser les écritures à partir des formes élémentaires qui les composent et qui peuvent être extraites grâce à un algorithme de segmentation de l'écriture. Ces caractéristiques permettent d'automatiser la tâche de vérification des scripteurs sans introduire de connaissances supplémentaires généralement utilisées par les experts (connaissance sur les mots ou sur les caractères). Les performances obtenues doivent encore être validées sur des tâches de vérification plus contraignantes. Nous pensons naturellement à la détection de faux. Mais en l'état actuel, l'approche permet d'ores et déjà de compléter les résultats que nous avons mis en évidence pour la tâche d'identification du scripteur [BEN 04], en constituant une méthode de rejet de scripteurs inconnus d'une base de référence.

Finalement cette étude met en évidence la capacité de l'écriture manuscrite à constituer un identifiant biométrique intéressant pour certains cas d'utilisation.

Références

[BEN 04] A. Bensefia, T. Paquet, L. Heutte. Documents manuscrits et recherche d'information. accepted for publication in Document Numerique, Hermès, 2004.

[BOV 03] Bovino, L. Impedovo S., Pirlo, G & Sarcinella, L. "Mutli-Expert Verification of Hand-Written Signatures". *Proceeding of the 7th International Conference on Document Analysis and Recognition*, ICDAR'03. Vol II, pp932-936. Scotland (2003)

[CHA 00] Cha, S.H., & Srihari, S. "Multiple Feature Integration for Writer Verification". *Proceedings of the 7th International Workshop on Frontiers in Handwriting Recognition; IWFHR VII* . pp 333-342 (2000).

[DUD 00] Duda R., Stork D., Hart P., Pattern Classification and Scene Analysis, Wiley & Sons, 2nd Edition, 2000.

[JAI 99] Jain, A.K, Bolle, R. & Pankanti, S. "Biometrics: Personal Identification in Network Society". Kluwer Academic Publishers, New York, 1999.

[JAR 99] Jarman, K.H., Hanlen, R.C., & Manzolillo, P.A. "Handwritting Examination : Moving From Art to Science". Pacific Northwest National Laboratory, 1999.

[MOR 00] Morris., R, N. « Forensic Handwritting Identification ». Academic Press, 2000.

[NOS 02] Nosary, A., Reconnaissance automatique de textes mansucrits par adaptation au scripteur, Thèse de doctorat de l'Université de Rouen, Janvier 2002.

[PAR 96] Parizeau, M. Ghazzali, N. & Hébert, J-F. « Codage Directionnel ave Attribut de Longueur et Optimisation de Matrice de Coût – Application en écriture cursive ». Actes 4ème Colloque National sur l'Écrit et le Document, CNED'96, pp ??- ??. 1996.

[SAP 90] Saporta G., « Probabilités analyse des données et statistiques », Edition Technip, (1990).

[SRI 01] Srihari., S., Cha., S., Arora., H., & Lee., S. "Individuality of Handwritten : A Validity Study". ICDAR'01. USA, 2001.

[WAG 74] Wagner, R. A. & Fisher, M. J. "The String to String Correction Problem". JACM vol, 21, N° 1, pp 168-173. 1974.

[ZIM 02] Zimmermann M., Bunke, H. " Automatic Segmentation of the IAM Off-line Handwritten {English} Text Database". 16th International Conference on Pattern Recognition, Canada, Vol 4, pp. 35-39, 2002.