



République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la
recherche scientifique

Université Larbi Tébessi - Tébessa



Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département : Mathématiques et Informatique

Mémoire de fin d'étude
Pour l'obtention du diplôme de *MASTER*
Domaine : Mathématiques et Informatique
Filière : Informatique
Option : Systèmes et Multimédias
Thème

*Contribution à l'identification des traits de personnalité
d'un scripteur à partir de documents manuscrits*

Présenté Par :
SMAILI SALAH EDDINE

Devant le jury :

<i>Mr. Abdeljalil Gattal</i>	<i>MAA</i>	<i>Université Larbi Tébessa</i>	<i>Président</i>
<i>Mr. Mohammed Aouine</i>	<i>MAB</i>	<i>Université Larbi Tébessa</i>	<i>Examineur</i>
<i>Mr. Chawki Djeddi</i>	<i>MCA</i>	<i>Université Larbi Tébessa</i>	<i>Encadreur</i>
<i>Mr. Mekhaznia tahar</i>	<i>MCB</i>	<i>Université Larbi Tébessa</i>	<i>Co-Encadreur</i>

Date de soutenance :/06/2019

Résumé

Dans ce manuscrit, un système automatique pour l'identification des traits de personnalité à partir de documents manuscrits est introduit. Le système proposé comprend trois étapes principales: la détection et l'élimination des lignes de portée, l'extraction de caractéristiques et la classification (identification de traits de personnalité). Dans la première étape, une technique pour la détection et l'élimination des lignes de portée présents dans les documents manuscrits utilisés est adaptée. Dans la deuxième étape, les distributions des longueurs de segments noirs et blancs, les distributions des directions des contours, les distributions des charnières des contours et les coefficients autorégressifs sont extraites à partir de documents manuscrits. Dans la troisième étape, nous avons utilisé les réseaux de neurones artificiels (ANN) pour la classification. Les expérimentations sont menées sur un ensemble de données qui comprend 453 documents manuscrits. Le taux d'identification le plus élevé est obtenu par l'utilisation des distributions des longueurs de segments blancs.

Keywords : Identification de traits de personnalité, documents manuscrits, longueurs de segments, directions des contours, charnières des contours, coefficients autorégressifs, réseaux de neurones artificiels.

Abstract

In this manuscript, a personality traits identification system based on handwritten documents is introduced. The proposed system consists of three main stages: staff lines detection and removal, feature extraction and classification (personality traits identification). In the first step, a method for the detection and elimination of staff lines present in the used handwritten documents is adapted. In the second step, run-lengths on black and white pixels, edge-hinge, edge direction and autoregressive features are extracted from handwritten documents. In the third step, we have used Artificial Neural Networks (ANN) as classifier. The experiments are conducted on a database that includes 453 handwritten documents. The highest identification rate is achieved by the use of black run-lengths features.

Keywords: Personality traits identification, handwritten documents, run-lengths, edge-hinge, edge-direction, autoregressive coefficients, artificial neural network.

في هذا العمل، تم تقديم نظام للتعرف على ملامح الشخصية باستعمال وثائق مكتوبة بخط اليد. يتكون النظام المقترح من ثلاث مراحل أساسية: كشف وإزالة الخطوط، استخراج المميزات وتصنيفها (التعرف على ملامح الشخصية). في الخطوة الأولى، تم تكييف طريقة للكشف عن الخطوط الموجودة في الوثائق المكتوبة بخط اليد وإزالتها. في الخطوة الثانية، يتم استخراج توزيع قياسات أطوال المسارات، توزيع مفصلات الحافة، توزيع اتجاه الحافة وكذلك معاملات الانحدار الذاتي من الوثائق المكتوبة بخط اليد. خلال المرحلة الثالثة استخدمنا مصنفات الشبكات العصبية الاصطناعية للتصنيف. أجريت التجارب على قاعدة بيانات تحتوي على 543 وثيقة مكتوبة بخط اليد. تم تحقيق أعلى دقة من خلال استخدام ميزات توزيع قياسات أطوال المسارات البيضاء.

الملخص

الكلمات المفتاحية : التعرف على الخطوط، مخطوطات تاريخية، أطوال المسارات، مفصلات الحافة، اتجاه الحافة، معاملات الانحدار الذاتي، الشبكات العصبية الاصطناعية.

DEDICACES

A ceux que j'aime ...et ceux qui m'aiment

Remerciements

Je remercie en premier Allah le tout puissant de m'avoir accordé la volonté et le courage pour réaliser ce mémoire.

Je souhait adresser mes vifs remerciements à ma chère famille : ma mère, mon père ainsi que mes sœurs, mon frère et lhaj Swissa pour leur soutiens, patience et générosité.

Je tiens de remercier Dr. Djeddi mon encadreur pour ces orientations, conseils judicieuses et surtout pour la patience.

Enfin, j'adresse mes remerciements à tous ce qui ont contribué de près ou de loin à la réalisation de ce travail.

Table des matières

1. Contexte du travail	8
2. Objectifs du travail	10
3. Organisation du mémoire	11
Chapitre 01:Analyse et classification automatique d'images de documents manuscrits	
1.1. Introduction.....	12
1.2. Authentification de signature	13
1.3. Reconnaissance de l'écriture manuscrite:.....	14
1.4. Identification de scripteurs.....	16
1.5. Identification des traits de personnalité.....	18
1.5.1. Travaux connexes	20
1.5.2. Modèle des cinq grands facteurs de personnalité	21
1.5.3. Caractéristiques de l'état de l'art.....	22
1.5.3.1. Zones	22
1.5.3.2. Ligne de base	23
1.5.3.3. Inclinaison de l'écriture	23
1.5.3.4. Taille de l'écriture	24
1.5.3.5. Pression du stylo.....	25
1.5.3.6. Espaces inter-mots	25
1.5.3.7. Marge.....	26
1.5.4. Processus d'identification des traits de personnalité... 26	
1.5.4.1. Acquisition des données.....	26
1.5.4.2. Prétraitements.....	26
1.5.4.3. Segmentation.....	26
1.5.4.4. Extraction de caractéristiques	27
1.5.4.5. Classification.....	27
1.6. Conclusion	27

Chapitre 02: Un système automatique pour l'identification de traits de personnalité

2.1. Introduction.....	28
2.2. Base de données	30
2.3. Détection et d'élimination des lignes de portée.....	31
2.3.1. Caractéristiques des lignes de portée	33
2.3.2. Recherche et élimination des lignes de portée	33
2.4. Extraction de caractéristiques	36
2.4.1. Distribution de longueurs de segments	36
2.4.2. Distribution de directions des contours	38
2.4.3. Distribution des charnières des contours.....	38
2.4.4. Coefficients autorégressifs	39
2.5. Classification.....	40
2.6. Résultats et discussion	41
2.7. Conclusion	45
BIBLIOGRAPHIE.....	47

Liste des figures

Figure 1.1.	Types d'écritures	15
Figure 1.2.	Les différentes zones de caractères	23
Figure 1.3.	Les différents types lignes de base	23
Figure 1.4.	Les différents types d'inclinaisons	24
Figure 1.5.	Les différents types d'écritures (petite taille, taille moyenne et grande taille).	22
Figure 1.6.	Les différents types de pression durant l'écriture	25
Figure 1.7.	Les différents types d'espaces inter-mots	25
Figure 2.1.	Exemples de deux pairs d'essais : images et texte, gauche et droite, respectivement.....	31
Figure 2.2.	Graphes du plus court chemin.	32
Figure 2.3.	La technique de détection d'une ligne de portée.....	34
Figure 2.4.	Résultat de la technique d'élimination des lignes de portée (a) Image originale, (b) Image après l'élimination des lignes de portée.	35
Figure 2.5.	Calcul de la matrice de longueur de segments, (a) Une image 8×6 avec deux valeurs de couleurs (0 et 1), (b) matrice de longueur de segments pour la direction 45°, (c) matrice de longueur de segments pour la direction 90°, (d) matrice de longueur de segments pour la direction 135°, (e) matrice de longueur de segments pour la direction 180°	37
Figure 2.6.	Les caractéristiques des directions des contours.....	38
Figure 2.7.	Les caractéristiques des charnières des contours.....	39
Figure 2.8.	Les trois couches d'un réseau de neurones artificiels (ANN)	41
Figure 2.9.	Répartition des documents de la base de données utilisée sur les ensembles d'apprentissage, de validation et de test.....	42
Figure 2.10.	Meilleurs résultats enregistrés en utilisant toutes les caractéristiques pour les quatre scénarios enregistrés.....	44

Liste des tableaux

Tableau 2.1: Répartition des scripteurs par trait de personnalité.	31
Tableau 2.2. Une vue d'ensemble des caractéristiques implémentées et leurs dimensionnalités.	40
Tableau 2.3. Les cinq caractéristiques pour le scénario 1.....	42
Tableau 2.4. Les cinq caractéristiques pour le scénario 2.....	43
Tableau 2.5. Les cinq caractéristiques pour le scénario 3.....	43
Tableau 2.6. Les cinq caractéristiques pour le scénario 4.....	43



INTRODUCTION GENERALE

1. Contexte du travail

L'écriture manuscrite représente l'un des modes de communication les plus anciens de notre civilisation qui ne cesse de se développer à travers les siècles. Elle nous permet de représenter nos pensées et de faire des échanges avec nos semblables [08]. Avec le développement des sciences, l'Homme a découvert que l'écriture manuscrite détient des informations variées qui caractérisent la personne, ses émotions, ses attributs démographiques et ses traits de personnalité, d'où l'utilisation de l'analyse et la reconnaissance automatique de documents manuscrits pour extraire des caractéristiques à l'aide de différentes méthodes, le choix de la méthode à utiliser dépend du type de données recherchées et du domaine étudié [08, 09].

La reconnaissance automatique de l'écriture manuscrite consiste à la création de systèmes automatiques capables de reconnaître des caractères, mots et textes manuscrits. Son principe est de transformer un texte écrit en une représentation compréhensible et reproductible par la machine. Malgré les grands progrès qu'a connus ce domaine ces dernières années, il demeure encore très actif vu la grande variabilité des écritures manuscrites et leurs différentes caractéristiques qui changent d'une personne à une autre [01].

L'identification du scripteur consiste à déterminer l'identité de l'auteur d'un échantillon d'écriture manuscrite à partir d'un ensemble de scripteurs [02]. La vérification du scripteur, c'est la tâche qui consiste à déterminer si deux documents manuscrits ont été écrits par la même personne ou par deux personnes différentes [10]. Si un n'importe quel texte peut être utilisé pour déterminer l'identité du scripteur, la tâche d'identification est indépendante du texte. Sinon, si un scripteur doit écrire un texte prédéfini

particulier pour s'identifier, la tâche d'identification est dépendante du texte. L'identification du scripteur peut être effectuée en-ligne, lorsque les informations temporelles et spatiales sur l'écriture sont disponibles, ou hors-ligne, si, l'identification est effectuée à partir d'une image numérisée de l'écriture manuscrite [10].

L'écriture peut être aussi un moyen efficace pour l'identification du genre d'un individu à partir de ses traces écrites [08], cette tâche est étroitement liée à celle de l'identification du scripteur, la différence étant que l'identification du scripteur est un problème à N classes alors que la classification des écritures selon le genre est un problème à deux classes. Les caractéristiques qui ont été effectivement appliquées à l'identification de scripteurs sont donc également susceptibles de bien fonctionner pour l'identification du genre [08].

Plusieurs études récentes ont montré que l'analyse de l'écriture manuscrite peut constituer un outil précieux pour la prédiction de la survenue de certaines maladies dégénératives (telle que la maladie de Parkinson) à un stade très précoce [12]. Un certain nombre d'évaluations cliniques sont menées pour diagnostiquer ces maladies, mais la plupart de ces tests ne sont efficaces que lorsque la maladie est à un stade relativement avancé [12]. Des attributs de l'écriture tels que la vitesse d'écriture, le temps d'écriture, la pression du stylet, les temps d'écriture, ainsi que les variations de la taille des lettres, etc. peuvent être calculés à partir d'échantillons de textes manuscrits issus de patients et de sujets témoins [13, 14, 15]. Ces attributs et leurs versions transformées peuvent être utilisés pour l'entraînement d'un ou plusieurs classifieurs afin d'apprendre à discriminer entre les deux classes. Au cours de la phase de validation, les caractéristiques extraites des échantillons en question seront transmises au classifieur formé, qui générera un score de confiance du sujet atteint/No atteint [09, 16]. Les échantillons rédigés peuvent être basés sur des modèles standards suggérés par les neuropsychologues [09, 16]. Ces modèles incluent normalement des spirales de dessin et l'écriture répétitive des alphabets avec des boucles. Les systèmes développés peuvent également intégrer à une tablette de numérisation de sorte que, outre les échantillons d'écriture stockés, le système puisse également acquérir des échantillons de nouveaux sujets et les classer dans l'une des deux classes.

L'écriture manuscrite a fait l'objet récemment de certaines études ayant comme objectif principal d'étudier ses liens avec l'activité cérébrale et les

aspects psychologiques de l'homme [11]. L'étude psychologique de l'écriture manuscrite dans le but de déterminer les traits de personnalité, les états psychologiques, le tempérament ou le comportement du scripteur est appelée graphologie et reste un domaine discutable dans la mesure où elle manque de norme, la plupart des interprétations de l'écriture manuscrite étant effectuées subjectivement par des personnes expérimentées (graphologues) [11]. Cependant, divers travaux de recherche ont montré le lien entre l'écriture manuscrite et les aspects neurologiques de l'homme, une de ces études étant celle de Plamondon [03], où il a été montré que le cerveau forme des caractères basés sur les habitudes du scripteur et chaque modèle cérébral neurologique forme un mouvement neuromusculaire distinct qui est similaire pour les individus ayant le même type de personnalité. Par conséquent, l'écriture manuscrite est, de ce point de vue, un miroir fidèle du cerveau de l'être humain.

Les graphologues analysent actuellement plusieurs caractéristiques de l'écriture afin d'évaluer les aspects psychologiques du scripteur, tels que l'épaisseur des traits, la trajectoire de l'écriture, la manière dont la lettre «t» ou «y» est écrite, ainsi que d'autres caractéristiques liées à la façon dont les lettres ou les mots sont écrits ou à la manière dont le texte est positionné sur la page [04, 05, 06].

2. Objectifs du travail

L'objectif essentiel de ce travail consiste à introduire une nouvelle technique indépendante du texte pour l'identification des traits de personnalité d'un individu à partir d'images scannées de son écriture manuscrite. La technique proposée exploite les distributions de longueurs de segments (Run-length), les distributions des directions des contours (Edge-direction), les coefficients autorégressifs ainsi que les distributions des charnières des contours (Edge-hinge) comme caractéristiques. La classification est effectuée à l'aide des réseaux de neurones (ANN). Les expérimentations seront réalisées sur la base de données fournie par les organisateurs de la Compétition "2018 ICPR Multimedia Information Processing for Personality & Social Networks Analysis Challenge" [07].

La technique proposée offre une alternative attrayante au questionnaire standard (Five Factor Model — FMM) [17] ou aux entretiens psychologiques actuellement utilisés pour évaluer la personnalité, car elle est plus facile à utiliser, demande moins d'efforts, et est plus rapide, plus efficace et enlève aussi la subjectivité, tant du côté du sujet que du

clinicien (comme les psychologues examinent généralement les résultats du questionnaire et partagent des opinions sur la personnalité de l'individu, opinions qui peuvent parfois être sujettes à des biais de sorte que différents psychologues puissent fournir des évaluations différentes).

3. Organisation du mémoire

Ce mémoire est structuré en deux chapitres. Le premier est consacré à la présentation des principaux concepts, outils et travaux relatifs à l'étude entreprise. Dans le deuxième chapitre du mémoire, nous abordons de manière détaillée nos choix conceptuels, la mise en œuvre ainsi que les résultats obtenus par le système proposé pour l'identification des traits de personnalité à partir de documents manuscrits hors-ligne.

CHAPITRE

1

Analyse et classification automatique d'images de documents manuscrits

Ce chapitre est consacré à la présentation des quelques domaines de recherche liés directement avec le problème étudié. Il présente les travaux connexes dans le domaine de l'identification des traits de personnalité à partir d'images de textes manuscrits. Nous décrivons les cinq grands traits de personnalité considérés dans le cadre de ce mémoire ainsi qu'une étude détaillée sur caractéristiques de l'écriture manuscrite et la personnalité. Nous terminons ce chapitre la présentation du processus d'analyse de l'écriture manuscrite pour détecter la personnalité d'une personne en se basant sur des documents manuscrits.

1.1. Introduction

La croissance des domaines de l'intelligence artificielle et de la reconnaissance des formes est due en grande partie à l'un des problèmes très difficiles d'identification des écritures manuscrites [08, 22]. Identifier les traits de personnalité d'un scripteur à partir de son écriture manuscrite est extrêmement essentiel aujourd'hui en raison de l'immense développement de la technologie et de ses applications dans de nombreux domaines [09, 10]. L'identification des traits de personnalité d'un scripteur à partir de ses traces écrites s'applique dans de nombreux domaines, tels que le recrutement et la sélection du personnel, la détection des problèmes psychologiques, la détection des talents cachés et des problèmes de santé, l'expérience passée, la criminalistique et l'interaction homme-machine [23].

Il est évident que la tâche d'identification des traits de personnalité d'un scripteur est devenue plus importante de nos jours. De toute évidence, le nombre de chercheurs impliqués dans ce problème complexe augmente en raison de ces opportunités [09,10]. Ce problème est basé sur l'écriture manuscrite se heurte à un certain nombre de sous-problèmes, tels que la conception d'algorithmes permettant d'analyser les écritures manuscrites de différents individus, identifier les caractéristiques les plus pertinentes permettant de bien caractériser les différents styles d'écritures et enfin bien choisir les techniques de classification automatique les mieux adaptées à ce problème [08, 09, 10].

1.2. Authentification de signature

La signature manuscrite consiste à la manière dont un individu signe son nom pour être unique [19, 18]. C'est une biométrie comportementale qui change sur une période de temps comme elle peut être facilement influencée par les conditions physiques et émotionnelles de la personne. Bien que les signatures nécessitent un contact avec l'instrument d'écriture et un effort de l'utilisateur, elles ont été acceptées par les tribunaux et les banques en tant que méthode d'authentification [19, 18].

Les systèmes de vérification de signatures manuscrites visent à décider si la signature en question est effectivement de l'individu revendiqué. En d'autres termes, ils sont utilisés pour classer les signatures manuscrites comme authentiques ou falsifiées [51]. Les falsifications sont généralement classées en trois types: les falsifications aléatoires, simples et qualifiées. Dans le cas de falsifications aléatoires, le faussaire n'a aucune information sur l'utilisateur ou sa signature et utilise sa propre signature à la place de ce dernier. Dans ce cas, la falsification contient une signification sémantique différente des signatures authentiques de l'utilisateur, présentant une forme globale très différente.

Dans le cas de falsifications simples, le faussaire connaît le nom de l'utilisateur, mais pas de sa signature. Dans ce cas, la falsification peut présenter plus de similitudes avec la signature authentique, en particulier pour les utilisateurs qui signent avec leur nom complet ou une partie de celui-ci. Dans les falsifications qualifiées, le faussaire a accès à la fois pour le nom et la signature de l'utilisateur et pratique souvent l'imitation de sa signature. Il en résulte des falsifications qui ressemblent davantage à la signature authentique et qui sont donc plus difficiles à détecter.

Les systèmes d'authentification de signatures manuscrites peuvent fonctionner selon deux modes différentes : identification et vérification [02]. L'identification c'est la tâche qui consiste à déterminer le propriétaire de la signature manuscrite, tandis que la vérification permet de confirmer ou de refuser l'identité en question et décider de l'authenticité ou de la falsification de la signature. La comparaison des échantillons de signatures déjà pris à ceux stockés dans la base de données, est mise en application dans ce processus d'identification et de vérification. Quand il s'agit de l'identification, la comparaison est un processus un-contre-tous alors que pour la vérification, elle est un processus un-contre-un et le résultat est booléen (oui /non).

Les classifieurs utilisés pour la vérification de signatures manuscrites peuvent être classifiés en deux groupes: dépendant de l'utilisateur et indépendant de l'utilisateur [51]. Dans le premier cas, un modèle est entraîné pour chaque utilisateur, en utilisant les signatures authentiques de l'utilisateur, et des falsifications aléatoires (en utilisant des signatures authentiques d'autres utilisateurs). Pendant la phase d'exploitation, le modèle entraîné pour l'identité revendiquée est utilisé pour classer les signatures en question comme authentiques ou falsifiées. D'un autre côté, l'approche indépendante de l'utilisateur n'implique qu'un seul classificateur pour tous les utilisateurs. Dans ce cas, le système apprend à comparer la signature en question avec une référence. Pendant la phase de test, le modèle est utilisé pour comparer une signature en question avec des échantillons réels de référence de l'individu revendiqué, pour prendre une décision. Une façon courante de former les systèmes indépendant de l'utilisateur consiste à utiliser une représentation de dissimilarité, où les entrées des classificateurs sont des différences entre deux vecteurs caractéristiques: $(x_Q - x_R)$, avec une étiquette binaire indiquant si les deux signatures proviennent du même utilisateur ou non.

1.3. Reconnaissance de l'écriture manuscrite:

La reconnaissance de l'écriture manuscrite est la capacité d'un micro-ordinateur à recevoir et à interpréter des entrées manuscrites provenant de sources telles que des documents papier, des photographies, des écrans tactiles et d'autres périphériques [01]. L'image du texte écrit peut être détectée en mode hors-ligne à partir d'un morceau de papier numérisé à l'aide d'un scanner où les informations recueillies se présentent sous la forme d'une image discrète constituée d'un ensemble

de pixels. Dans le mode en-ligne, il s'agit de reconnaître l'écriture au fur et à mesure de son tracé. Le texte est saisi avec un stylo et une tablette à numériser, les informations recueillies sont constituées par une suite ordonnée de points (définis par leurs coordonnées) échantillonnées à cadence fixe où l'écriture prend l'aspect d'un couple de signaux temporels numérisés. La reconnaissance en-ligne se fait en temps réel, c'est à dire, pendant l'écriture. La reconnaissance en-ligne est généralement beaucoup plus efficace que la reconnaissance hors-ligne car les échantillons sont beaucoup plus informatifs. En revanche, elle nécessite un matériel beaucoup plus couteux et impose de fortes contraintes au scripteur puisque la capture de l'encre doit se faire au moment de la saisie (capture synchrone) et non a posteriori (capture asynchrone) comme les Smartphone et les Tablet PC. Ces systèmes sont utilisés dans plusieurs équipements électroniques [20].

La disposition spatiale du texte joue un rôle important dans les systèmes de reconnaissance de l'écriture manuscrite. Une écriture pré-casé ou guidée, comme les formulaires par exemple est plus aisée à la reconnaissance qu'une écriture cursive.

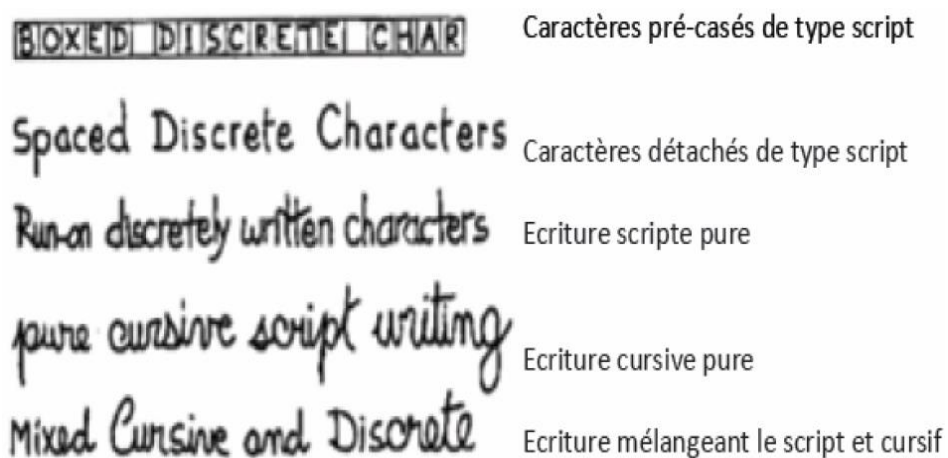


Figure 1.1 Types d'écritures [20]

La difficulté de reconnaissance croît avec le nombre de scripteurs. On peut distinguer trois types de systèmes de reconnaissance d'écriture, classés par ordre de complexité de fonctionnement :

- Systèmes mono-scripteur, où un seul scripteur peut utiliser le système de reconnaissance après apprentissage de son écriture;

- Systèmes multi-scripteur où le système peut reconnaître les écritures d'un groupe restreint d'individus, soit après adaptation à l'écriture de chacun, soit sans adaptation ;
- Systèmes omni-scripteur où le système est censé reconnaître toutes les écritures. Dans ce cas, la variabilité intra-scripteur s'ajoute à la variabilité inter-scripteur.

De la même manière que le nombre des scripteurs, la difficulté de la reconnaissance croît avec la taille du vocabulaire considéré. Trois grandes catégories d'applications peuvent être distinguées :

- Applications à vocabulaire très limité où le nombre de mots (ou de symboles) à reconnaître constitue un lexique de taille réduite (inférieure à 100 mots), comme par exemple dans le cas de l'ensemble des mots utilisés pour écrire les montants de chèques postaux ou bancaire. Il est alors possible de confronter chaque mot à reconnaître à l'ensemble des mots du lexique;
- Applications à vocabulaire étendu, mais pouvant être réduit de façon dynamique, comme l'ensemble des noms de rues associés à un bureau de poste distributeur;
- Applications à vocabulaire très étendu où plusieurs milliers ou dizaines de milliers de mots formant un dictionnaire et pour lesquels la confrontation systématique n'est plus possible comme par exemple le dictionnaire des noms de commune.

1.4. Identification de scripteurs

L'intérêt porté à l'écriture est grandissant, du fait de sa nature artistique intéressante mais également du fait de la variété d'informations qu'elle peut véhiculer. En effet, certains s'attachent à faire parler les écritures pour qu'elles révèlent les traits comportementaux de leurs auteurs ou simplement leurs identités [08,21].

L'écriture manuscrite est considérée comme un acte personnel : chaque scripteur est caractérisé par son écriture, par la reproduction de détails et d'habitudes inconscientes. C'est pourquoi dans certains cas d'expertise l'analyse des échantillons d'écriture a même valeur que l'analyse des empreintes digitales. En effet que l'authentification d'une personne à partir de son écriture manuscrite représente un domaine important de la

biométrie, surtout lorsque l'écriture représente l'unique trace exploitable laissée par les personnes [08,21].

L'écriture désigne tout media conventionnel utilisé par des groupes d'individus pour communiquer entre eux. C'est une représentation du langage humain sous une forme graphique. Bien qu'elle puisse différer d'une population à une autre, ses concepts de base restent pratiquement les mêmes (alphabet, mots, phrases) [08]. L'écriture est facilement quantifiable, son acquisition peut se faire soit en mode dynamique (en ligne) à l'aide d'un stylo électronique et d'une tablette d'acquisition numérique, soit en mode statique ou différé (hors-ligne) où son acquisition se fait au moyen d'un scanner ou d'une caméra [08, 10, 21].

Telle la parole ou la démarche, l'écriture fait partie des connaissances permanentes d'un individu qui, une fois apprises, ne disparaissent jamais, sauf dans certains cas très rares. La propriété centrale qui permettrait de rattacher l'écriture au groupe des identifiants biométriques est l'individualité. Autrement dit, est-ce que l'écriture manuscrite est suffisamment individuelle pour discriminer un individu par rapport à toute une population? Il existe deux méthodes d'analyses qui permettent de valider une telle propriété : une analyse manuelle, effectuée par des experts en écritures, ou bien une analyse automatique dans laquelle il faut concevoir un système capable de caractériser les écritures manuscrites de chaque scripteur.

Plusieurs travaux ont abordé cette question et ont plutôt conclu à l'individualité de l'écriture. Ces travaux postulent que c'est une caractéristique comportementale d'auto-expression, agissant sous l'influence de plusieurs facteurs propres, éducationnels, culturels, intentionnels, émotionnels, circonstanciels et contextuels. Ces facteurs sont le fruit d'un ensemble de propriétés:

- Le degré de maturité graphique de l'auteur (enfant, adulte, ..),
- La force d'expression graphique,
- L'état psychologique du scripteur (fatigue, stress, ..),
- Le système d'écriture appris,
- Les connaissances linguistiques,
- Le matériel utilisé pour écrire (papier, encre, surface, ..),
- Le développement musculaire du scripteur,

Un système d'identification du scripteur a pour principale fonction de fournir une liste ordonnée des scripteurs de sa base de référence selon leur degré de similitude avec l'échantillon d'écriture en présence. Et dans l'approche d'identification, l'utilisateur dispose d'un seul document, et cherche à identifier son auteur parmi un ensemble de N scripteurs connus. Tandis que l'approche d'identification nécessite une décision du type plus proche voisin [08, 22].

Bien que l'identification du scripteur s'inscrive dans la même problématique que la reconnaissance de l'écriture, elle ne semble pas poser le même type de difficultés cependant. En effet, la tâche d'identification peut tirer profit de la variabilité des écritures afin de les discriminer, tandis que la tâche de reconnaissance doit au contraire parvenir à s'affranchir de la variabilité entre les scripteurs pour identifier le message textuel quel qu'en soit le scripteur. La variabilité de l'écriture d'un scripteur dépend de facteurs psychologiques multiples qu'il est difficile de déterminer avec précision [08, 21, 22].

1.5. Identification des traits de personnalité

L'analyse de l'écriture manuscrite [24, 26] est une méthode scientifique pour reconnaître, évaluer et comprendre la personnalité du scripteur à travers une analyse de la forme des mots extraient de son écriture manuscrite. L'écriture peut être utilisée manuscrite pour reconnaître vraie personnalité d'un individu, y compris son comportement, ses charges émotionnelles, son estime de soi, sa colère, son imagination, sa peur, son honnêteté et bien d'autres traits de personnalité.

Les experts légaux de l'écriture manuscrite qui identifient la personnalité à travers les échantillons d'écriture manuscrite sont appelés graphologues. L'état psychologique ainsi que comportement d'un individu peuvent être identifiés à l'aide de la voix (parole), les expressions faciales, les gestes, la posture et la façon de s'habiller. Souvent, le style extérieur reflète le style intérieur. L'écriture est aussi une des façons expressives qui parle de la nature, l'état psychologique et le comportement d'un individu et plus précisément d'un scripteur.

Comme avancé dans les sections précédentes, l'écriture manuscrite [24, 25] est unique pour chaque individu. Elle doit être identique et unique pour un scripteur, quel que soit le moyen utilisé pour la production de l'écriture. L'écriture manuscrite est écrite par le cerveau, pas par la main.

Ainsi, l'écriture manuscrite est également appelée comme écriture cérébrale. Chaque trait de personnalité a un schéma neurologique dans le cerveau humain. Chaque schéma neurologique de cerveau génère un mouvement neuromusculaire unique en son genre, identique pour chaque individu présentant ce trait de personnalité spécifique. Chaque mouvement en écriture révèle un trait de personnalité particulier. La graphologie est la science qui identifie ces traits dans l'écriture manuscrite et décrivant le trait de personnalité correspondant.

L'analyse de l'écriture manuscrite ou graphologie a de larges champs d'application dans les domaines tels que le recrutement, la psychologie, le diagnostic médical, la criminalistique et l'interaction homme-machine. L'écriture manuscrite représente la personnalité et le comportement des êtres humains et peut donc être utilisée pour le recrutement et la sélection du personnel. L'écriture manuscrite révèle également beaucoup de choses sur le scripteur, telles qu'un problème psychologique, la moralité, les talents cachés, les problèmes de santé, l'expérience passée, etc. L'analyse de l'écriture manuscrite peut être encore utilisée pour obtenir un aperçu du psychisme de la personne et elle sert également comme outil de diagnostic. L'écriture manuscrite révèle les conditions psychologiques et physiologiques du patient ; Il est donc utilisé comme outil d'évaluation dans le diagnostic médical et psychologique.

Auparavant, l'analyse de l'écriture manuscrite se faisait manuellement en consacrant beaucoup de temps à prédire la nature de la personne. En analyse manuelle, la précision de l'analyse dépend des compétences du graphologue. Le graphologue est également enclin à la fatigue lorsque plusieurs échantillons doivent être analysés. Pour obtenir les services d'un graphologue expérimenté, les coûts sont élevés. D'autre part, l'analyse automatisée de l'écriture manuscrite est très rapide, précise, très peu coûteuse et pratique dans la prédiction de la personnalité humaine.

Dans les sous sections suivantes, nous nous concentrons principalement sur les travaux connexes dans le domaine de l'identification des traits de personnalité à partir d'images de textes manuscrits, ensuite nous présentons une brève description des cinq grands traits de personnalité et puis nous expliquons les caractéristiques de l'écriture manuscrite et la personnalité et enfin nous décrivons le processus d'analyse de l'écriture manuscrite pour détecter la personnalité d'une personne en se basant sur des documents manuscrits.

1.5.1. Travaux connexes

L'identification des traits de la personnalité d'un individu à partir de ses traces écrites a fait l'objet de plusieurs travaux de recherche effectués par les psychologues à l'opposé de leurs homologues informaticiens qui n'ont pas accordé à ce problème l'importance qu'il mérite. Durant la dernière décennie, les informaticiens ont commencé à s'intéresser à ce domaine de recherche en proposant quelques systèmes automatiques pour l'automatisation de cette tâche.

Rahiman A.M et al. [24] ont proposé un outil d'analyse de l'écriture manuscrite hors ligne appelé "HABIT"; qui signifie "*Handwriting Analysis Based Individualistic Traits Prediction*" et qui est utilisé pour détecter les traits de personnalité d'un scripteur à partir des images scannées de son écriture manuscrite. Champa H N et al. [26] ont également proposé un système à base de règles dont les paramètres d'entrée sont la pression du stylo, la ligne de base, la hauteur de la lettre *t* et la forme de la boucle de la lettre *y*, Ce système prédit les traits de personnalité à partir de ces caractéristiques. La ligne de base a été trouvée en utilisant la méthode de polygonalisation, la pression du stylo a été trouvée en utilisant la valeur seuil du niveau de gris, la boucle de la lettre *y* est analysée par la technique de transformée de Hough.

Une autre méthode a été proposée par Mukherji S et al. Dans [30] où les auteurs ont utilisé différents algorithmes pour extraire des caractéristiques comme la ligne de base, l'inclinaison, la taille, la marge, etc. de l'image manuscrite d'un document pour l'identification de la personnalité. Dhadwal A. et al. ont présenté dans [31] un système utilisant les réseaux de neurones artificiels pour connaître le psychisme de la personne à travers un échantillon d'écriture manuscrite d'un individu. Des techniques de traitement d'image ont été utilisées pour la détection des sections sous pression, la détection des contours, la segmentation des lignes et la segmentation des caractères pour calculer les angles d'inclinaison, la hauteur et la pression à l'aide d'une technique de seuillage.

Toujours dans le même cadre, Joshi P et al. [32] ont proposé un outil d'apprentissage automatique utilisant un classifieur kNN avec apprentissage incrémentiel afin d'améliorer l'efficacité de l'analyse des écritures manuscrites. Les caractéristiques utilisées pour l'analyse de la personnalité incluent la détection de la position de ligne de base,

l'inclinaison des lettres, la hauteur de la barre de la lettre "t et la marge. La ligne de base a été calculée à l'aide de la méthode de polygonisation, la marge a été calculée à l'aide de la méthode de balayage vertical, la hauteur de la barre de la lettre t et l'inclinaison des mots ont été calculées à l'aide du modèle de correspondance.

Dans les travaux proposés par Mutalib S et al dans [33], les auteurs ont présenté un système pour déterminer le niveau de contrôle des émotions de scripteur à l'aide de l'inférence floue. L'émotion du scripteur a été analysée par la ligne de base de l'écriture et l'utilisation de l'inférence de Mamdani. Ce système a été principalement développé pour aider un directeur d'une entreprise afin de détecter l'état émotionnel de ses conseillers.

Grewal P. K et al. [34] ont proposé un système basé sur les réseaux de neurones artificiels pour la prédiction du comportement en utilisant des caractéristiques comme l'inclinaison, la ligne de base, la pression du stylo, la lettre "f" et la lettre "i". La ligne de base et l'inclinaison ont été analysées à l'aide de la méthode de polygonisation, la pression du stylo a été analysée en utilisant une valeur de seuil de niveau de gris, et la lettre "i" et la lettre "f" ont été analysées en utilisant la correspondance de modèles.

1.5.2. Modèle des cinq grands facteurs de personnalité

Le modèle des cinq grands facteurs de personnalité [17] est un modèle bien connu pour décrire la personnalité d'un individu. Il est basé sur cinq traits de personnalité de base, regroupés en sous-facteurs, comme suit:

- Extraversion: C'est un trait qui fait référence aux personnes qui expriment facilement des émotions positives. L'extraversion comprend des traits de type dynamique, loquace et assertif.
- Convivialité : c'est un trait qui fait référence aux personnes qui ont tendance à faire preuve de compassion au lieu de suspectes, ainsi que utiles et tempérées.
- Conscience professionnelle : c'est un trait qui fait référence aux personnes qui sont fiable, et qui ont une prédilection pour des comportements soigneusement planifiés et orientés vers des résultats et des réalisations.

- Stabilité émotionnelle: c'est un trait qui fait référence aux personnes qui manquent de stabilité émotionnelle et de contrôle et ont tendance à éprouver des sentiments négatifs, telles que la colère et l'anxiété, ainsi qu'une vulnérabilité à la dépression.
- Ouverture à l'expérience : c'est un trait qui fait référence aux personnes qui peuvent facilement exprimer leurs émotions et qui ont un désir d'aventure, d'appréciation de l'art et d'idées originales.

1.5.3. Caractéristiques de l'état de l'art

Notre recherche bibliographique nous a permis de constater que diverses caractéristiques de l'écriture manuscrite ont été utilisées pour la détection des traits de la personnalité d'un individu. Ces caractéristiques manuscrite incluent: Les zones, la ligne de base, la pression, la taille, l'espacement, l'inclinaison, la marge, etc. Une brève description de ces caractéristiques est présentée ci-dessous.

1.5.3.1. Zones

L'écriture manuscrite est divisée en trois parties appelées partie supérieure, partie centrale et partie inférieure, comme le montre la figure 1.2. La personnalité peut être identifiée en analysant ces trois zones. Les sections des trois zones servent à définir l'équilibre entre les trois principaux domaines de développement de l'ego du scripteur, la sphère intellectuelle et spirituelle de l'individu, les activités quotidiennes et les commandes intuitives inconscientes [38].

- La zone supérieure représente l'avenir, la conscience spirituelle, l'aspiration intellectuelle et culturelle, la perception mentale et le fantasme.
- La zone centrale représente le présent, l'expression réaliste, la pratique de l'ego et l'expression émotionnelle.
- La zone inférieure représente le passé, le mémoire, les pulsions inconscientes, la perception sensuelle, les pulsions de base, les pulsions inconscientes et les besoins biologiques.

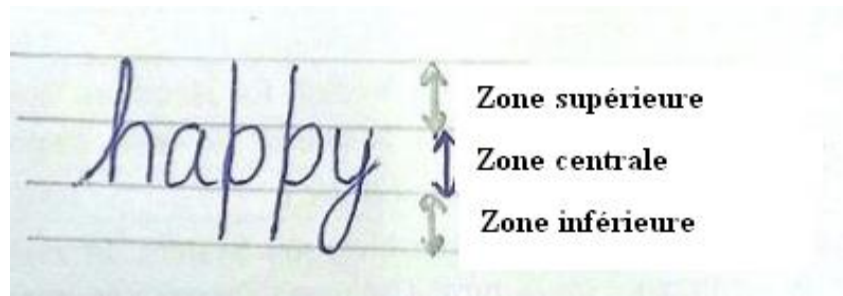


Figure 1.2. Les différentes zones de caractères [09].

1.5.3.2. Ligne de base

La ligne de base peut être ascendante, descendante, droite, comme illustrée à la figure 1.3. La ligne de base est utilisée pour trouver l'effort émotionnel et la nature du scripteur.

- Une ligne de base ascendante indique que la personne est optimiste, représente l'espoir et la gaieté, indique l'excitabilité le comportement colérique.
- Une ligne de base descendante indique que la personne est pessimiste, qu'elle est fatiguée mentalement temporairement et a des troubles digestifs.
- Une ligne de base droite montre que la personne a un comportement extérieur stable, qu'elle est droite, réaliste et disciplinée.

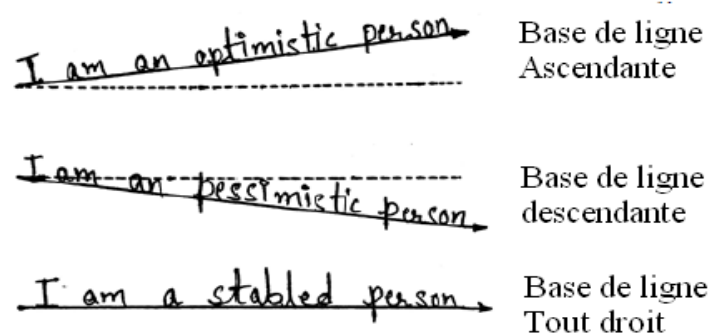


Figure 1.3. Les différents types lignes de base [09].

1.5.3.3. Inclinaison de l'écriture

L'inclinaison de l'écriture est définie par la direction de la pente de la lettre et est mesurée par l'angle formé entre la lettre et la ligne de base. L'inclinaison reflète les émotions, le degré de contrôle sentimental et du contrôle émotionnel du scripteur. Il existe trois types d'inclinaison, ces

types sont : l'inclinaison verticale, l'inclinaison vers la droite et l'inclinaison vers la gauche [09]. La figure 1.4 montres les différents types d'inclinaisons.

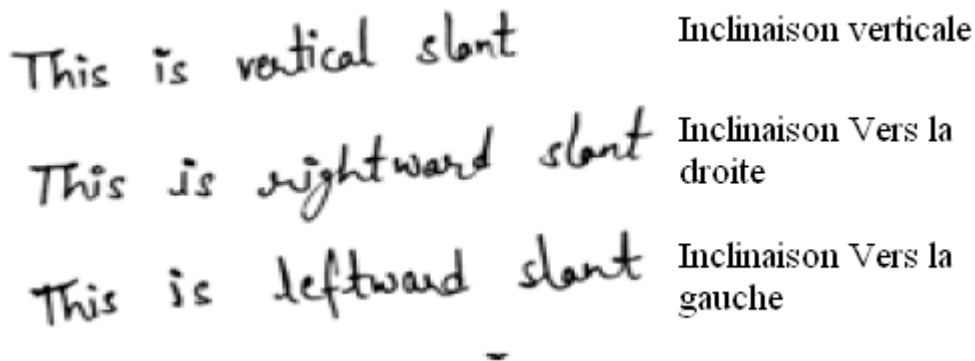


Figure 1.4. Les différents types d'inclinaisons [09].

1.5.3.4. Taille de l'écriture

La taille de l'écriture est mesurée par la hauteur verticale des lettres. La taille peut être grande, moyenne ou petite, comme le montre la figure 1.5. La taille nous indique l'importance que le scripteur accorde à lui-même et à ses propres actions. Elle indique également comment le scripteur se surprendre sur son environnement. Par exemple, le scripteur dont l'écriture est de grande taille aborde la vie avec extraversion et excès d'indulgence et le scripteur dont l'écriture est petite aborde la vie avec solitude et timidité [09].

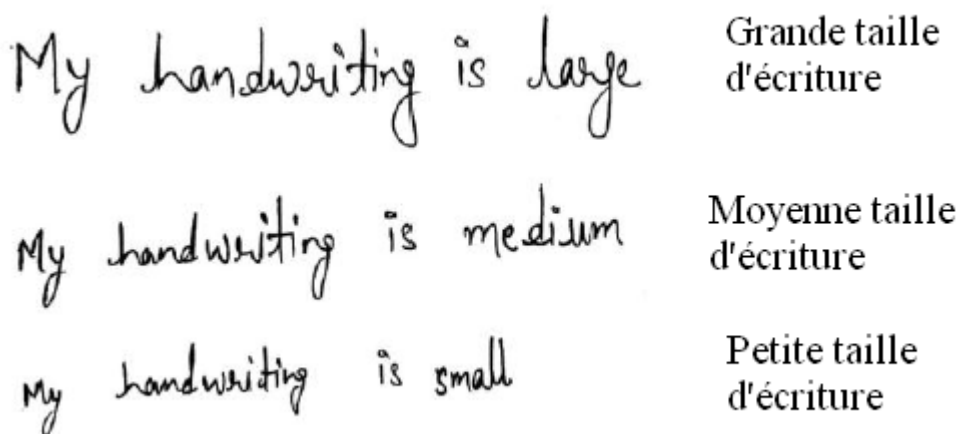


Figure 1.5. Les différents types d'écritures (petite taille, taille moyenne et grande taille) [09].

1.5.3.5. Pression du stylo

La force appliquée au moment de l'écriture est considérée comme une pression de stylo. Elle peut être lourde, légère ou moyenne comme le montre la figure 1.6. A partir de la pression de stylo, nous pouvons analyser l'énergie mentale de scripteur. Une forte pression indique que le scripteur est énergique, actif, anxieux, vigoureux, en colère, en alerte et ponctué. Une pression moyenne montre que le sentiment du scripteur n'est pas très intense. Une légère pression indique le calme, la passivité, le manque d'énergie et la maladie [25].

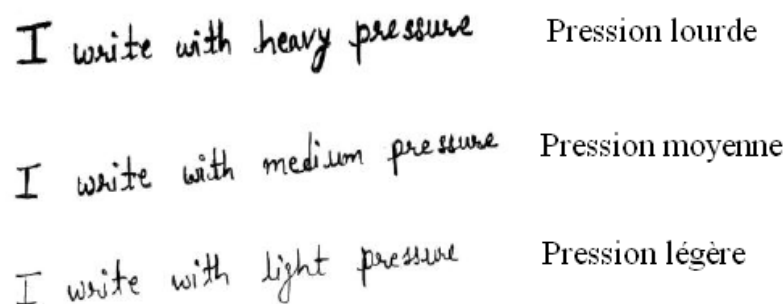


Figure 1.6. Les différents types de pression durant l'écriture [09].

1.5.3.6. Espaces inter-mots

L'espace inter-mots (voir figure 1.7) représente la distance que le scripteur souhaite maintenir entre lui et les autres personnes, sa proximité avec les autres et son intelligence. Un large espace indique la discrimination, l'indépendance, le bon goût, l'exclusivité, le snobisme, la fierté et la capacité d'organiser son travail. Un petit espace indique l'incapacité d'être seul, le mauvais goût, l'amabilité et l'intrusion.

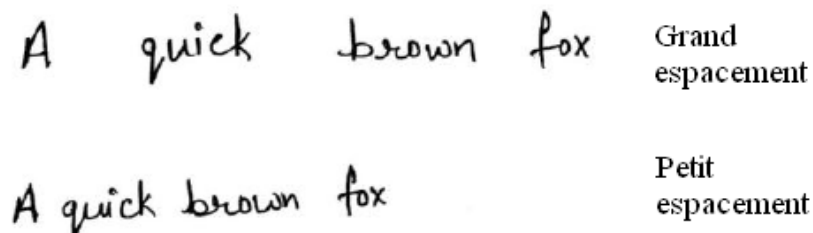


Figure 1.7. Les différents types d'espaces inter-mots [09].

1.5.3.7. Marge

La marge est la quantité d'espace laissée par le scripteur à gauche ou à droite, en haut ou en bas de la page [36]. Quand le scripteur commence à écrire sur une feuille de papier vierge, il prend un certain espace de marge sur cette feuille de papier et après cela commence à écrire. La marge est de différents types, c'est-à-dire la marge supérieure, inférieure, gauche et droite. La marge peut définir le passé et l'avenir, les ajustements, l'intelligence, la rapidité et la véracité [25].

1.5.4. Processus d'identification des traits de personnalité

Pour identifier la personnalité d'un individu à partir de son écriture manuscrite, plusieurs étapes sont nécessaires, ces étapes peuvent inclure : l'acquisition de données, les prétraitements, la segmentation, l'extraction des caractéristiques et la classification (identification) [09].

1.5.4.1. Acquisition des données

Les échantillons d'écriture sont prélevés sur un papier blanc de format A4. Cet échantillon d'écriture manuscrite sert comme entrée dans le système. L'échantillon d'écriture manuscrite est numérisé à l'aide d'un scanner et stocké sous forme de fichier image (JPEG ou PNG, etc...) ou il peut être pris avec un appareil photo.

1.5.4.2. Prétraitements

Les prétraitements sont effectués pour améliorer la qualité de l'image en entrée et la qualité de certaines caractéristiques de l'image est également améliorée pour un traitement ultérieur. Les prétraitements qui peuvent être appliqués à l'image comprennent l'élimination du bruit, la binarisation et la normalisation. La binarisation convertit l'image en niveaux de gris en une image binaire. Les techniques de suppression du bruit sont appliquées pour supprimer les données indésirables et améliorer la qualité de l'image. La normalisation est utilisée pour supprimer certaines variations des styles d'écriture manuscrite et pour simplifier les formes des symboles.

1.5.4.3. Segmentation

La segmentation est le processus permettant de segmenter la page manuscrite en lignes, mots et caractères. La segmentation en lignes est

utilisée pour segmenter un paragraphe en un ensemble de lignes ensuite les lignes obtenues sont segmentées en mots et ces derniers doivent être segmentés en caractères.

1.5.4.4. Extraction de caractéristiques

L'extraction de caractéristiques est le processus qui consiste à extraire les attributs les plus pertinents à partir d'une image de texte dans l'objectif de caractériser son scripteur. Ces caractéristiques peuvent être la taille, la ligne de base, l'inclinaison, la marge, les zones, etc.

1.5.4.5. Classification

La classification est utilisée pour reconnaître les traits de personnalité du scripteur d'un document manuscrit. Les différentes caractéristiques extraites agissent comme une entrée pour le classifieur.

1.6. Conclusion

Dans ce chapitre, nous avons présenté dans un premier temps, quelques domaines de recherche qui sont en liaison directe avec le problème étudié, ces domaines incluent : l'authentification de signatures manuscrite, la reconnaissance de l'écriture manuscrite et l'identification de scripteurs à partir de documents manuscrits, nous nous sommes concentré ensuite sur les travaux connexes dans le domaine de l'identification des traits de personnalité à partir d'images de textes manuscrits, puis, nous avons décrit les cinq grands traits de personnalité considérés dans le cadre de ce mémoire. Après avoir présenté une étude détaillé sur caractéristiques de l'écriture manuscrite et la personnalité. Nous avons terminé ce chapitre par une conclusion précédée par la présentation du processus d'analyse de l'écriture manuscrite pour détecter la personnalité d'une personne en se basant sur des documents manuscrits.

CHAPITRE

2

Un système automatique pour l'identification de traits de personnalité

Ce chapitre présente notre contribution principale qui consiste en une étude sur l'identification de traits de personnalité à partir de documents manuscrits. La méthode proposée est basée sur l'extraction d'un ensemble de caractéristiques texturales à partir d'échantillons de documents manuscrits et l'entraînement d'un classifieur afin qu'il puisse identifier les traits de personnalité du scripteur d'un document en question. Des attributs comme l'orientation, la courbure et la texture sont estimés en calculant les distributions de longueurs de segments (Run-length) [21], les distributions des directions des contours (Edge-direction) [21], les distributions des charnières des contours (Edge-hinge) [21] et les coefficients autorégressifs [08]. La classification est effectuée à l'aide des réseaux de neurones artificiels (ANN). La méthode proposée a été évaluée en utilisant une base de données contenant 543 documents manuscrits où des résultats intéressants ont été enregistrés.

2.1. Introduction

Les dernières années ont connu une augmentation significative de la recherche dans différents domaines de l'analyse et la classification de documents manuscrits. Des avancées notables dans ce domaine ont abouti à de nombreux travaux incluant la reconnaissance de l'écriture manuscrite [20], la reconnaissance de scripteurs [21], la classification des écritures manuscrites selon le sexe du scripteur [08], la détection prématurée de certaines maladies dégénératives [12] et la classification de manuscrits anciens [40]. Ces travaux sont tous basés sur l'analyse de

documents manuscrits anciens ou modernes afin d'extraire certains attributs discriminatoires permettant leur classification.

Malgré le développement considérable dans le domaine de l'analyse et la classification de documents manuscrits, l'identification de traits de personnalité à partir de documents manuscrits est restée parmi les mécanismes d'authentification les moins étudiés par les chercheurs du domaine. L'identification automatique de traits de personnalité est à notre avis un problème intéressant de reconnaissance des formes [09] et les travaux présentant des tentatives timides pour résoudre ce problème ont été résumés dans un certain nombre d'études récentes [41].

L'identification de traits de personnalité consiste à identifier le trait de caractère d'un scripteur parmi un ensemble de plusieurs candidats. Le système doit alors avoir appris au préalable des échantillons d'écriture de chaque trait de personnalité. Bien que l'identification des traits de personnalité s'inscrive dans la même problématique que la reconnaissance de l'écriture, elle ne semble pas poser le même type de difficultés. En effet, la tâche d'identification de traits de personnalité doit tirer profit de la variabilité des écritures afin de les discriminer, tandis que la tâche de reconnaissance doit, au contraire, parvenir à s'affranchir de cette variabilité entre les scripteurs pour identifier le message textuel quel qu'en soit le scripteur.

La méthode proposée offre une alternative attrayante au questionnaire standard (Five Factor Model — FMM) [17] ou aux entretiens psychologiques actuellement utilisés pour évaluer la personnalité, car elle est plus facile à utiliser, demande moins d'efforts, et est plus rapide, plus efficace et enlève aussi la subjectivité, tant du côté du sujet que du clinicien (comme les psychologues examinent généralement les résultats du questionnaire et partagent des opinions sur la personnalité de l'individu, opinions qui peuvent parfois être sujettes à des biais de sorte que différents psychologues puissent fournir des évaluations différentes).

Ce chapitre est organisé comme suit. Dans la section suivante, nous décrivons la base de données utilisée. Nous présentons ensuite la description de la méthode de détection et d'élimination des lignes de portée choisie ainsi que la technique de classification utilisée dans les sections 2.3 et 2.4 respectivement, suivies par la description des caractéristiques proposées dans la section 2.5. Les résultats

expérimentaux et leur analyse sont présentés dans la section 2.6, tandis que la dernière section conclut ce chapitre.

2.2. Base de données

La méthode proposée est évaluée la base de données fournie par les organisateurs d'une compétition sur l'identification des traits de personnalité à partir documents manuscrits qui a été organisée en conjonction de la conférence internationale sur la reconnaissance des formes, ICPR 2018 [45]. Le corpus utilisé dans le cadre de cette compétition est composé d'essais écrits par des d'étudiants mexicains de premier cycle universitaire. Pour chaque essai, deux fichiers sont disponibles: une transcription manuelle du texte et une image numérisée de la feuille originale où le sujet a rédigé l'essai à la main.

Les textes des transcriptions manuelles ont été étiquetés pour marquer certains phénomènes manuscrits, à savoir: FO (le mot précédent est mal orthographié), D (il y a un dessin ici; par exemple une signature), <IN> (insertion de mots dans le texte), <MD> (modification d'un mot, c'est-à-dire une correction d'un mot; par exemple. quand le sujet a oublié d'écrire une lettre et modifier le mot), <DL> (élimination d'un mot), <NS> (quand deux mots ont été écrits ensemble; par exemple. I am instead of I am) and, SB (syllabification). Chaque essai est étiqueté avec cinq classes correspondant aux cinq grands traits de personnalité. Ces traits sont : l'extraversion, la convivialité, la conscience professionnelle, la stabilité émotionnelle et l'ouverture à l'expérience.

Le nombre total d'échantillons dans le corpus utilisé est de 543, il est divisé en trois sous-ensembles: apprentissage, validation et test. L'ensemble de validation est utilisé pour le réglage de paramètres du système si nécessaire, tandis que le sous ensemble de test est utilisé pour l'évaluation des performances du système proposé. Le tableau 2.1 montre la répartition des scripteurs (également appelés sujets) par trait de personnalité. Enfin, un exemple d'un couple image/texte est donné dans le Figure 2.1.

Trait	Nombre d'échantillons
Extraversion	253
convivialité	258
Conscience professionnelle	291
Stabilité émotionnelle	231
Ouverture à l'expérience	245

Tableau 2.1: Répartition des scripteurs par trait de personnalité.

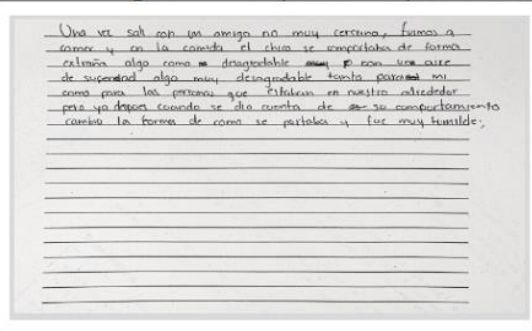
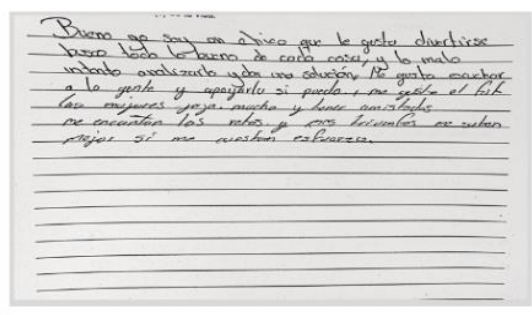
	<p><ANR_0021 e:6.0 a:5.5 c:5.0 s:4.5 g:5.5> Una vez sali <FO:sali> con un amigo no muy cercano, fuimos a comer y en la comida el chico se comportaba de forma extraña algo como <DL> desagradable <DL> <DL> con un <MD> aire de superioridad <MD> algo muy desagradable tanto para <DL> mí <FO:mí> como para las personas que estaban en nuestro alrededor pero ya despues <FO:después> cuando se dio cuenta de <DL> su comportamiento <u>cambió</u> <FO:cambió> la forma de como <FO:cómo> se portaba y fue muy humilde. </ANR_0021></p>
	<p><ANR_0028 e:6.0 a:5.5 c:6.5 s:4.5 g:6.5> Bueno soy un chico que le gusta divertirse busco todo lo bueno de cada cosa, y lo malo intento analizarlo y <NS> dar una solución. Me gusta escuchar a la gente y apoyarla si puedo , no me gusta el fut <FO:futbol> las mujeres jaja. mucho y tener amistades me encantan los retos y mis triunfos me saben mejor si me cuestan esfuerzo. </ANR_0028></p>

Figure 2.1 : Exemples de deux pairs d'essais : images et texte, gauche et droite, respectivement. [42].

2.3. Détection et d'élimination des lignes de portée

La détection des lignes de portée "Staff lines" est une étape très importante dans notre travail. Sa grande importance réside dans la facilité avec laquelle nous pouvons ensuite procéder à l'extraction des caractéristiques à partir des textes manuscrits. Cette étape est généralement réalisée en utilisant une technique de binarisation des images de documents manuscrits en fixant des seuils qui peuvent être locaux ou globaux. Cependant, la technique de binarisation choisie peut entraîner la perte de certaines informations pertinentes de l'image du

document manuscrit et introduire du bruit dans les images ce qui peut influencer considérablement sur les étapes ultérieures du processus d'identification étudié. Il est donc nécessaire de bien choisir une méthode qui minimise la perte d'information due à la binarisation.

Pour la détection et l'élimination des lignes de portée présentes dans les images de la base de données utilisée pour l'évaluation des performances de la méthode proposé, nous avons utilisé la technique de détection et d'élimination des lignes de portée proposée dans [50] qui considère l'image d'un document manuscrit comme un graphe (voir Figure 2.2) où les pixels représentent les nœuds et les arcs réfèrent aux pixels voisins. Chaque nœud p est doté d'une valeur qui correspond à son poids et définie par le degré de la nuance dont il est doté. Bien entendu, le poids de chaque arc $w(v_1, v_2)$, reliant les nœuds v_p et v_q est une fonction des valeurs des pixels correspondants. Le chemin (v_1, v_n) reliant deux nœuds distants est une liste unique de nœuds v_1, v_2, \dots, v_n avec (v_i, v_j) tout couple de nœuds adjacents appartenant à l'ensemble $\{v_1, v_n\}$. Le coût du chemin est égal à la somme des coûts de chaque arc reliant deux nœuds adjacents du dit chemin :

$$w(v_1, v_n) = \sum_{i=2}^n w(v_{i-1}, v_i)$$

Un chemin reliant deux nœuds donnés est dit le plus court (shortest path) si son cout est le plus bas parmi les chemins reliant les mêmes nœuds.

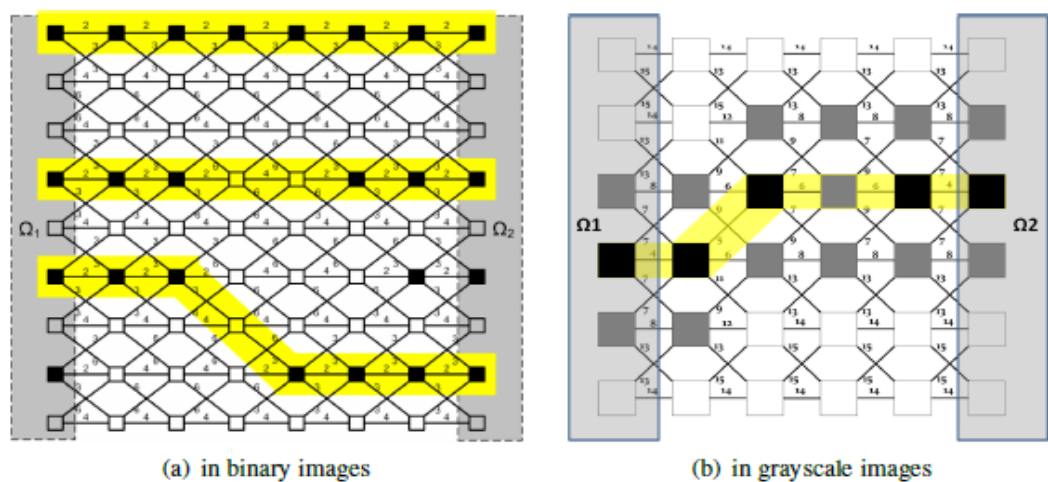


Figure 2.2. Graphes du plus court chemin [43].

Certaines techniques de prétraitements sont nécessaires pour effectuer les opérations de détection et d'élimination de lignes de portée. Ces techniques consistent à calculer la distance et la hauteur des lignes de portée et puis à les éliminer. Une brève description de ces deux techniques est donnée dans les sous sections suivante.

2.3.1. Caractéristiques des lignes de portée

Une ligne de portée est un sous graphe constitué de pixels d'un coût élevé (nuance noire) construite sous une forme quasi rectiligne (absence de zigzag). Elle relie généralement les limites gauche et droite d'une image manuscrite. Les pixels constituant une ligne de portée, appartenant à l'ensemble situé dans la limite d'un angle de 45° pour chaque pixel à connectivité 8. Une ligne de portée peut être extraite à partir du chemin le plus court reliant les sous-graphes situant à la limite gauche et droite de l'image à raison d'un et un seul pixel à chaque colonne de l'image.

2.3.2. Recherche et élimination des lignes de portée

Soit une image I de dimension $N1 \times N2$ pixels, une ligne de portée est en général définie sous la forme:

$$s = \{(x, y(x))\}_{x=1}^{N1} \forall x |y(x) - y(x-1)| \leq 1 \text{ avec } y : [1, \dots, N1] \rightarrow [1, \dots, N2]$$

Son coût peut être calculé de la manière suivante:

$$C(s) = \sum_{i=2}^{N1} w(v_{i-1}, v_i).$$

La ligne minimisant le coût C peut être définie à l'aide d'un algorithme de programmation dynamique qui consiste à parcourir l'image I à partir de la deuxième colonne de gauche en direction du bord droit en cumulant les coûts respectifs de tous les chemins partant d'un pixel $p(i,j)$ se trouvant sur la première colonne à gauche, on aura donc:

$$c(i, j) = \min \begin{cases} c(i-1, j-1) + w(p_{i-1, j-1}; p_{i, j}) \\ c(i-1, j) + w(p_{i-1, j}; p_{i, j}) \\ c(i-1, j+1) + w(p_{i-1, j+1}; p_{i, j}) \end{cases}$$

Avec $w(p_{i,j}, p_{l,m})$ dénote le poids du bord incident avec les pixels aux positions (i,j) et (l,m). La ligne de portée est donc égale à $\min_{j \in \{1, \dots, N2\}} c(N1, j)$

Pour déterminer d'autres lignes, il fallait auparavant effacer chaque ligne trouvée pour ne pas la comptabiliser une seconde fois. En générale, et pour chaque pixel d'une ligne, on prend celui se trouvant à droite à 0° . Si ce n'est pas le cas (absence de nuance adéquate), on prend celui à 45° en haut ou bas. Si la ligne n'est pas horizontale ou comportant des zigzags, on peut se permettre de prendre des pixels à 45° plusieurs fois à la limite d'une valeur I . Si ça dépasse le seuil défini, on se contente de reprendre des pixels de nuance inférieure à la limite de l'horizontale ou d'ignorer la recherche, car la ligne est peut-être n'est pas de portée. La figure 2.3 illustre le processus de détection de ligne de portée alors que la figure 2.4 présente un échantillon de texte manuscrit avant et après l'élimination des lignes de portée.



Figure 2.3 : La technique de détection d'une ligne de portée.

A la edad de 18 años ingreso a la Universidad Autónoma de la Ciudad de México en la licenciatura de Filosofía e Historia de la Ideas. Por motivo de una invalidez más viable, decidí realizar mi examen para ingresar a la UAM-C, esta universidad me queda muy cerca de mi hogar. En esta ingreso a los 20 años en la carrera de Humanidades, me encuentro muy agusto y satisfecho por dos motivos principalmente. En primer lugar porque me di cuenta que soy capaz de obtener un abstracción que me proponga (lograr entrar a la UAM en mi primer intento). En segundo lugar porque la UAM me ha ofrecido múltiples opciones para que me sienta muy cómodo dentro de ella, un ejemplo claro es la diversidad de Servicios que ofrece y la flexibilidad en ellos. En la actualidad, al estar cursando mi primer trimestre estoy dispuesto a realizar lo suficiente para llevar acaba en tiempo y forma mi historial académico y todo lo aprendido en esta carrera después lo podré llevarlo a la práctica en la sociedad, y como, pues mi intención principal de ser licenciado en Humanidades es para poder ser docente y junto a ello seguir preparándome en torno en la Historia Mexicana.

(a)

A la edad de 18 años ingreso a la Universidad Autónoma de la Ciudad de México en la licenciatura de Filosofía e Historia de la Ideas. Por motivo de una invalidez más viable, decidí realizar mi examen para ingresar a la UAM-C, esta universidad me queda muy cerca de mi hogar. En esta ingreso a los 20 años en la carrera de Humanidades, me encuentro muy agusto y satisfecho por dos motivos principalmente. En primer lugar porque me di cuenta que soy capaz de obtener un abstracción que me proponga (lograr entrar a la UAM en mi primer intento). En segundo lugar porque la UAM me ha ofrecido múltiples opciones para que me sienta muy cómodo dentro de ella, un ejemplo claro es la diversidad de Servicios que ofrece y la flexibilidad en ellos. En la actualidad, al estar cursando mi primer trimestre estoy dispuesto a realizar lo suficiente para llevar acaba en tiempo y forma mi historial académico y todo lo aprendido en esta carrera después lo podré llevarlo a la práctica en la sociedad, y como, pues mi intención principal de ser licenciado en Humanidades es para poder ser docente y junto a ello seguir preparándome en torno en la Historia Mexicana.

(b)

Figure 2.4. Résultat de la technique d'élimination des lignes de portée (a) Image originale, (b) Image après l'élimination des lignes de portée.

2.4. Extraction de caractéristiques

Le système d'identification de traits de personnalité proposé est basé sur un ensemble de caractéristiques texturales extraites à partir des images de documents manuscrits. Ces caractéristiques sont décrites dans les sous sections suivantes.

2.4.1. Distribution de longueurs de segments

Nous caractérisons les différents documents en calculant la distribution des longueurs de segments (Run-length) [21, 47]. Ces caractéristiques sont déterminées à partir d'une image binaire du document manuscrit considérée, où les pixels noirs correspondent à la trace d'encre et les pixels blancs correspondent à l'arrière-plan. Les longueurs de segments sont calculées directement à partir de l'image complète du document. Afin de calculer les longueurs de segments, l'image est balayée dans les quatre directions principales: horizontale, verticale, diagonale gauche et diagonale droite. L'histogramme normalisé de ces longueurs de segments est interprété comme une probabilité de distribution afin de caractériser le document.

Nous commençons par la définition d'un «run» comme un ensemble de pixels consécutifs ayant des niveaux de gris similaires, dans une direction donnée. Si $A_i A_j$ est un segment composé de pixels $A_i, A_{i+1}, \dots, A_{j-1}, A_j$ de couleurs identiques, le pixel A_{i-1} doit être de couleur différente de celle du pixel A_i , le pixel A_j doit être aussi de couleur différente de celle du pixel A_{j+1} .

Nous définissons ensuite une méthode statistique de caractérisation de la texture qui consiste à effectuer le comptage du nombre de segments de pixels de même intensité dans une direction donnée et représenter les résultats dans une matrice appelée matrice de longueur de segments P . Pour cela, une direction ($0^\circ, 45^\circ, 90^\circ$ ou 135°) et un nombre de niveaux de gris sont préalablement fixés. La valeur contenue dans la case (l, n) de la matrice est égale au nombre de segments de longueur l et de niveaux de gris n . Donc le nombre de colonnes de la matrice est dynamique car elle dépend de la longueur du plus long segment. De par sa conception, ce calcul est symétrique, il est par conséquent inutile de la calculer dans les quatre directions complémentaires ($180^\circ, 225^\circ, 270^\circ$ ou 315°) sachant qu'on considère ici huit directions possibles entre le pixel étudié et ses

voisins). Ainsi, en général, la matrice est établie pour les quatre directions principales (0° , 45° , 90° et 135°).

La méthode d'extraction de caractéristiques proposée est illustrée par un exemple dans la figure 2.5. Nous considérons une image de 8×6 avec deux couleurs $C = \{0, 1\}$. Cette image représente le chiffre '2'.

0	1	1	1	1	0
1	0	0	0	0	1
1	0	0	0	0	1
0	0	0	0	1	0
0	0	0	1	0	0
0	0	1	0	0	0
0	1	0	0	0	1
1	1	1	1	1	1

(a)

2	1	2	2	1	1
9	2	0	0	0	1

(b)

3	2	3	2	1	0	0	0
11	4	0	0	0	0	0	0

(c)

7	3	2	1	0	1
15	2	0	0	0	0

(d)

4	2	3	3	0	0
9	0	0	1	0	1

(e)

Figure 2.5. Calcul de la matrice de longueur de segments, (a) Une image 8×6 avec deux valeurs de couleurs (0 et 1), (b) matrice de longueur de segments pour la direction 45° , (c) matrice de longueur de segments pour la direction 90° , (d) matrice de longueur de segments pour la direction 135° , (e) matrice de longueur de segments pour la direction 180° [21]

Chaque élément de la matrice indique le nombre de fois où l'image contient un segment d'une longueur $\{1, 2, 3, 4, 5, 6\}$ dans les directions 45° , 135° et 180° et des segments d'une longueur de $\{1, 2, 3, 4, 5, 6, 7, 8\}$ dans la direction 90° . Le premier élément de la première ligne de la matrice correspond au nombre de fois où la couleur 0 apparaît isolément, le deuxième élément est le nombre de fois où la couleur 0 apparaît en segments de deux pixels et ainsi de suite... La ligne suivante capte les mêmes informations à partir de l'image pour la couleur 1.

2.4.2. Distribution de directions des contours

Cette distribution directionnelle est calculée très rapidement en utilisant la représentation du contour, avec l'avantage supplémentaire que l'influence de la largeur de trace d'encre est éliminée. La distribution des directions du contour est extraite en considérant l'orientation des fragments locaux du contour. Un fragment est déterminé par deux pixels de contour (x_k, y_k) et $(x_{k+\omega}, y_{k+\omega})$ pris à une certaine distance ω . L'angle que fait le fragment avec l'horizontale est calculé en utilisant:

$$\varphi = \arctan\left(\frac{y_{k+\omega} - y_k}{x_{k+\omega} - x_k}\right)$$

Lorsque l'algorithme s'exécute sur le contour, l'histogramme des angles est construit. Cet histogramme d'angle est ensuite normalisé à une distribution de probabilité qui donne la probabilité de trouver dans l'image de document manuscrit un fragment de contour orienté à chaque φ . L'angle φ réside dans les deux premiers quadrants parce que, sans informations en ligne, nous ne savons pas avec quelle inclination l'utilisateur a signé. L'histogramme est étalé sur l'intervalle 0° - 180° et divisé en n sections (cases). Par conséquent, chaque section pèse 15° , ce qui rend les caractéristiques proposées suffisamment robustes [21, 46]. La figure 2.6 présente une description graphique des caractéristiques des directions des contours.

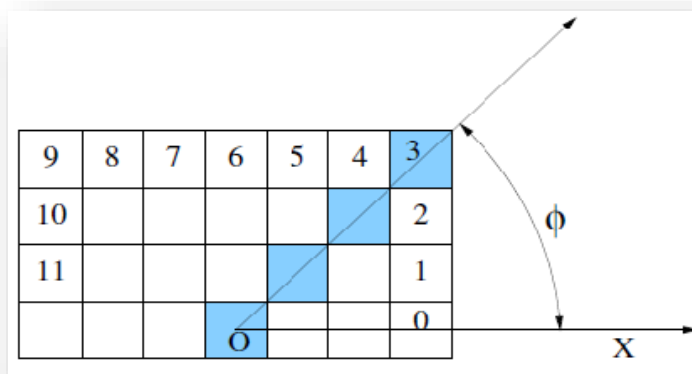


Figure 2.6. Les caractéristiques des directions des contours. [21].

2.4.3. Distribution des charnières des contours

Afin de capturer la courbure du contour, ainsi que son orientation, les distributions des charnières des contours sont utilisées. Cette méthode

d'extraction de caractéristiques [21, 46] est similaire à celle décrite précédemment, mais elle est un peu plus complexe. L'idée centrale est de considérer dans le voisinage, non pas un, mais deux fragments de bord émergeant du pixel central et de calculer ensuite la distribution de probabilité conjointe des orientations des deux fragments. La figure 2.7 présente une description graphique des caractéristiques des charnières des contours.

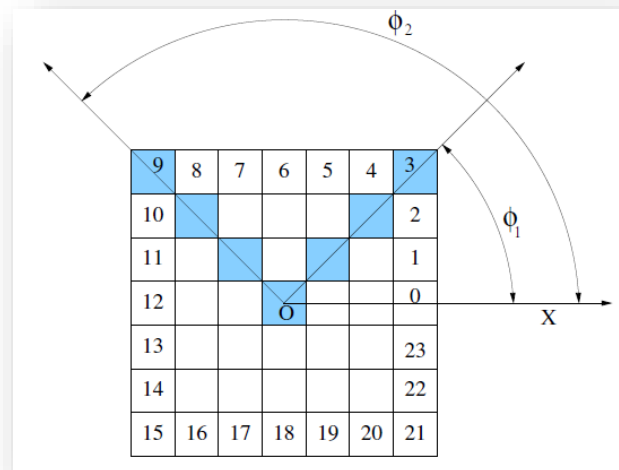


Figure 2.7. Les caractéristiques des charnières des contours. [21].

2.4.4. Coefficients autorégressifs

Un modèle non causal de modèle bidimensionnel, appelé modèle autorégressif bidimensionnel (AR), est également utilisé pour la caractérisation de scripteurs. La motivation derrière l'utilisation d'un tel modèle provient de son application réussie dans des domaines tels que l'analyse de texture [48, 49]. Aucune connaissance spécifique à un script n'est requise pour utiliser ce modèle, qui fournit un cadre viable pour traiter les scripteurs écrivant dans plusieurs scripts.

Si le modèle AR est appliqué à chaque pixel d'une image, la mise en œuvre peut nécessiter un calcul important, même pour une image de taille raisonnable. Par conséquent, au lieu de calculer le modèle AR dans tous les pixels des images binaires, nous calculons le modèle uniquement en pixels noirs. Cela réduit considérablement les exigences de calcul. Il est à noter que dans les images numérisées de texte manuscrit, environ 3% des pixels ou moins sont en noir.

Le tableau 2.2 résume, pour chacune des caractéristiques utilisées, le numéro correspondant, la description et la dimension.

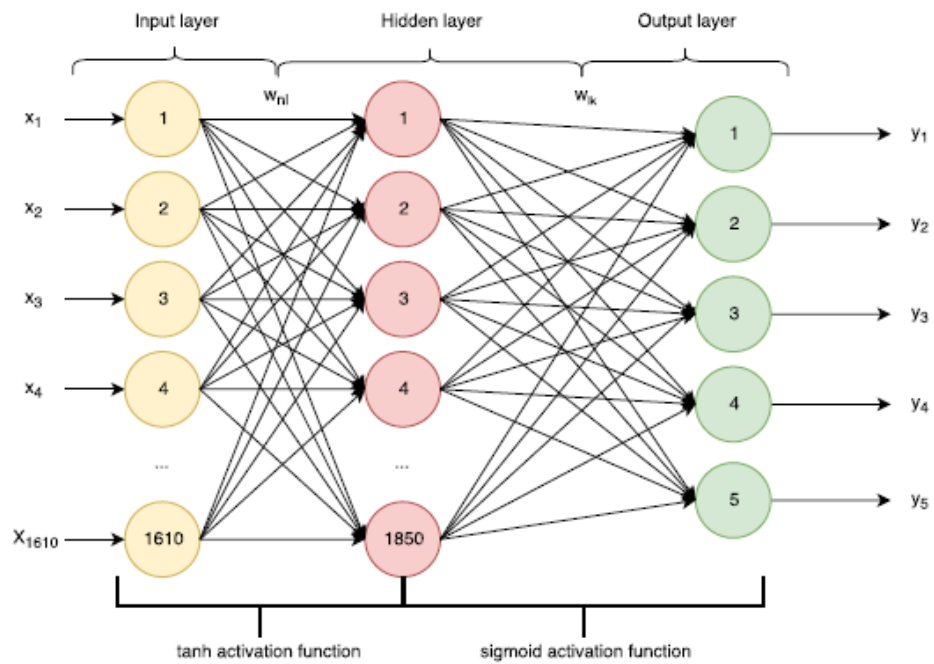
Caractéristique	Description	Dimension
f_1	Distribution des charnières des contours avec un fragment de 7 pixels	3136
f_2	Distribution des longueurs de segments noirs	400
f_3	Coefficients autorégressifs	24
f_4	Distribution des directions des contours	40
f_5	Distribution des longueurs de segments blancs	400

Tableau 2.2. Une vue d'ensemble des caractéristiques implémentées et leurs dimensionnalités.

2.5. Classification

Une fois que les images de documents à comparer sont représentées par leurs caractéristiques, nous procédons à l'utilisation de ces vecteurs de caractéristiques pour la classification. L'entraînement et la classification sont effectués à l'aide des réseaux de neurones artificiels (ANN) [08]. En utilisant les ensembles de données d'apprentissage et de test tandis que les différents paramètres du classifieur sont déterminés empiriquement sur l'ensemble des données de validation. Une brève introduction à ce classifieur est donnée dans les paragraphes suivants.

Le réseau de neurones artificiels (ANN) est un réseau à trois couches, la couche d'entrée ayant le même nombre de neurones que la dimension d'un ensemble particulier de caractéristiques, la couche de sortie comprend cinq neurones correspondant aux cinq traits de personnalités étudiés dans le cadre de ce mémoire (extraversion, convivialité, conscience professionnelle, stabilité émotionnelle et ouverture à l'expérience), alors que le nombre de neurones dans la couche cachée est déterminé en fonction de la dimension du vecteur de caractéristiques (à l'aide l'ensemble de données de validation). La figure 2.8 présente une description graphique des trois couches d'un réseau de neurones artificiels (ANN).



La figure 2.8 : Les trois couches d'un réseau de neurones artificiels (ANN) [44].

2.6. Résultats et discussion

Nous présentons et analysons les performances des caractéristiques proposées dans l'identification des traits de personnalité. La mesure de performances utilisée est le taux d'identification. Toutes les expérimentations sont menées sur la base de données présentée dans la section 2.2. Nous avons envisagé quatre scénarios d'évaluation différents en changeant les tailles des bases d'apprentissage (d'entraînement), de validation et de test pour chaque scénario. La figure 2.9 présentent La répartition des documents de la base de données utilisée sur les ensembles d'apprentissage, de validation et de test.

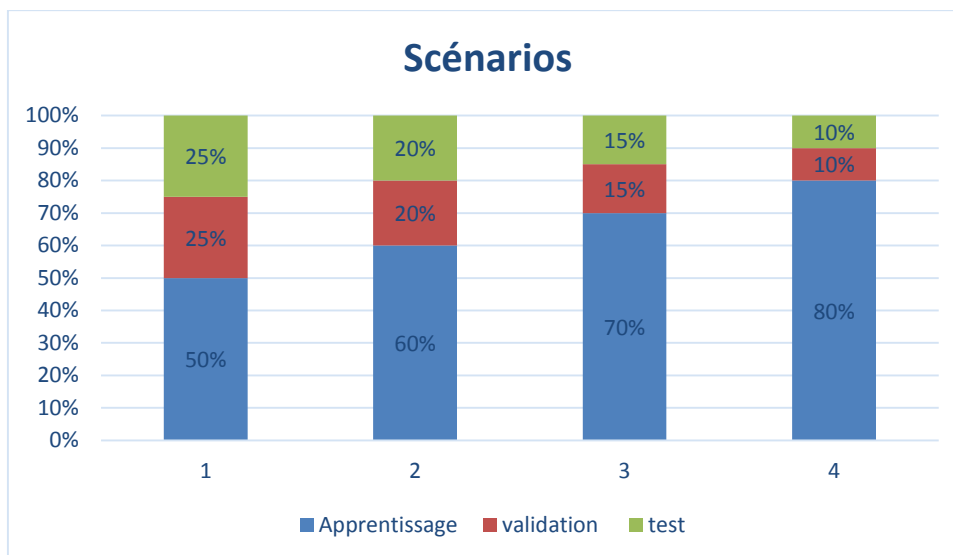


Figure 2.9. Répartition des documents de la base de données utilisée sur les ensembles d'apprentissage, de validation et de test.

Nous avons effectué plusieurs séries d'expérimentation dans l'objectif de déterminer les meilleurs paramètres du classifieur utilisé conduisant aux meilleurs performances. A cet effet, nous avons varié le nombre de neurones dans la couche cachée de 100 à 5000 neurones.

Les tableaux 2.3, 2.4, 2.5 et 2.6 présentent les résultats réalisés en utilisant les cinq types de caractéristiques pour les quatre scénarios considérés. La figure 2.2 montre les meilleurs résultats enregistrés en utilisant toutes les caractéristiques pour les quatre scénarios considérés.

Caractéristique	Nombre de neurones dans la couche cachée				
	2400	2500	2600	2700	2800
<i>f1</i>	62,98%	61,03%	64,13%	59,67%	59,74%
<i>f2</i>	57,35%	53,81%	53,74%	50,13%	57,20%
<i>f3</i>	58,38%	52,27%	60,48%	49,58%	57,94%
<i>f4</i>	57,16%	57,53%	59,85%	60,04%	51,93%
<i>f5</i>	60,41%	52,85%	67,99%	57,72%	60,63%

Tableau 2.3. Résultats réalisés avec les cinq caractéristiques pour le scénario 1.

Caractéristique	Nombre de neurones dans la couche cachée				
	3300	3400	3500	3600	3700
<i>f1</i>	61,20%	57,60%	67,80%	71,00%	54,70%
<i>f2</i>	53,90%	51,70%	54,00%	56,40%	59,40%
<i>f3</i>	54,70%	59,00%	54,60%	60,20%	58,20%
<i>f4</i>	57,00%	53,40%	56,40%	55,00%	56,50%
<i>f5</i>	71,90%	59,70%	59,80%	48,60%	59,20%

Tableau 2.4. Résultats réalisés avec les cinq caractéristiques pour le scénario 2.

Caractéristique	Nombre de neurones dans la couche cachée				
	4000	4100	4200	4300	4400
<i>f1</i>	60,07%	61,47%	59,67%	53,48%	62,25%
<i>f2</i>	53,15%	50,90%	54,11%	51,20%	49,65%
<i>f3</i>	52,63%	51,90%	52,12%	58,60%	54,25%
<i>f4</i>	51,75%	50,87%	53,33%	57,20%	52,27%
<i>f5</i>	62,87%	59,71%	73,66%	57,13%	49,43%

Tableau 2.5. Résultats réalisés avec les cinq caractéristiques pour le scénario 3.

Caractéristique	Nombre de neurones dans la couche cachée				
	500	600	700	800	900
<i>f1</i>	61,25%	54,07%	53,85%	50,94%	51,57%
<i>f2</i>	54,36%	55,06%	59,52%	54,44%	50,35%
<i>f3</i>	54,62%	60,96%	57,94%	52,60%	57,35%
<i>f4</i>	63,68%	57,31%	52,74%	55,95%	52,60%
<i>f5</i>	57,75%	53,96%	73,08%	53,44%	51,93%

Tableau 2.6. Résultats réalisés avec les cinq caractéristiques pour le scénario 4.

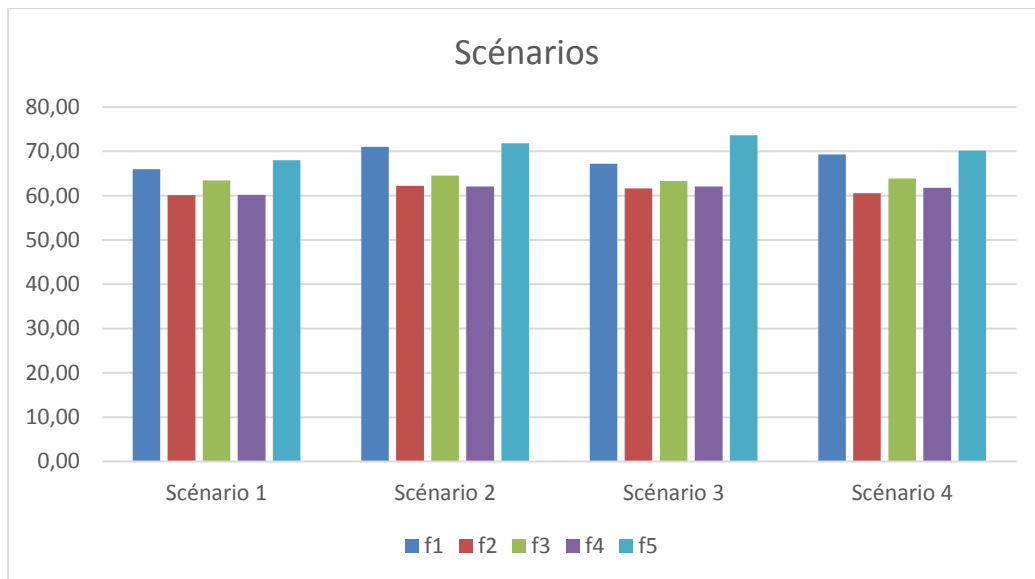


Figure 2.10. Meilleurs résultats enregistrés en utilisant toutes les caractéristiques pour les quatre scénarios enregistrés.

Une analyse des résultats présentés dans la figure 2.10 conduit aux observations suivantes :

- La méthode de distributions des longueurs de segments blancs (f5) donne des performances meilleures que celles enregistrées par toutes les autres caractéristiques. Elle a donné les meilleures performances dans tous les scénarios.
- Il convient de noter également que les longueurs de segments blancs (f5) sont plus informatives que les longueurs de segments noirs (f2).
- Le scénario 2 donne des performances meilleures que celles enregistrées par tous les autres scénarios pour les distributions des charnières des contours (f1), les distributions des longueurs de segments noirs (f2), les coefficients autorégressifs (f3) et les distributions des directions des contours (f4).
- Le scénario 3 donne les meilleures performances avec les distributions des longueurs de segments blancs (f5).
- Enfin, les résultats obtenus démontrent clairement le potentiel des caractéristiques proposées pour l'identification des traits de personnalité à partir d'images scannées de l'écriture manuscrites.

2.7. Conclusion

Ce travail avait pour objectif de présenter une méthode pour l'identification des traits de personnalité d'un scripteur à partir de ses traces écrites. Nous avons utilisé un ensemble de caractéristiques texturales qui ont montré des résultats prometteurs sur une base de données de documents manuscrits. Les évaluations ont été effectuées sur une base de données contenant 543 échantillons de textes manuscrits. Les résultats obtenus pour l'identification de traits de personnalité sont encourageants. Ils reflètent l'efficacité des caractéristiques texturales sur des documents manuscrits. Dans tous les cas, certaines caractéristiques se sont montrées plus performantes que d'autres caractéristiques.

La contribution que nous avons proposée dans le cadre de ce mémoire nous ont permis d'aboutir à des résultats prometteurs, mais nous ont aussi ouvert plusieurs voies pouvant être exploitées dans le futur. Les études ultérieures sur ce sujet seront destinées à introduire des caractéristiques supplémentaires et ensuite appliquer un mécanisme de sélection de caractéristiques pour savoir quelles sont les caractéristiques les plus discriminantes pour ce problème et pour des problèmes similaires. Il est nécessaire de rappeler que la performance de la méthode proposée ne dépend pas seulement de la technique de classification utilisée, mais aussi des caractéristiques choisies.

Dans ce cadre, il serait très intéressant d'exploiter la combinaison de caractéristiques proposées afin d'améliorer les performances de la méthode proposée. Pour la technique de classification utilisée, nous pensons qu'il serait intéressant d'envisager l'utilisation d'autres techniques de classification que celle que nous avons adoptée dans le présent mémoire. Il serait aussi très intéressant aussi d'envisager et d'expérimenter des possibilités de combinaison de techniques de classification, d'évaluer d'autres techniques de détection et d'élimination des lignes de portée "staff lines" et de considérer des bases de données plus volumineuses que celle utilisée dans le cadre du présent travail.

CONCLUSION

GENERALE

Ce travail a abordé le problème de l'identification des traits de personnalité d'un individu à partir de ses traces écrites en utilisant les distributions de longueurs de segments (Run-length), les distributions des directions des contours (Edge-direction) et les distributions des charnières des contours (Edge-hinge) ainsi que les coefficients autorégressifs comme caractéristiques. Les réseaux de neurones artificiels (ANN) ont été employés pour la classification. La méthode proposée a été évaluée en utilisant une base de données contenant 543 documents manuscrits où des performances très encourageantes ont été réalisées en particulier lors de l'utilisation des distributions des longueurs de segments blancs.

Les études ultérieures sur ce sujet seront destinées à introduire des caractéristiques supplémentaires et ensuite appliquer un mécanisme de sélection de caractéristiques pour savoir quelles sont les caractéristiques les plus discriminantes pour ce problème et pour des problèmes similaires. Il est nécessaire de rappeler que la performance du système proposé ne dépend pas seulement de la technique de classification utilisée, mais aussi des caractéristiques choisies.

Dans ce cadre, il serait très intéressant d'exploiter la combinaison de caractéristiques proposées dans ce mémoire afin d'améliorer les performances de la méthode proposée. Pour la technique de classification utilisée, nous pensons qu'il serait intéressant d'envisager l'utilisation d'autres techniques de classification que celle que nous avons adoptée dans le présent mémoire. Il serait aussi très intéressant aussi d'envisager et d'expérimenter des possibilités de combinaison de techniques de classification, d'évaluer d'autres techniques de détection et d'élimination des lignes de portée "staff lines" et de considérer des bases de données plus volumineuses que celle utilisée dans le cadre du présent travail.

BIBLIOGRAPHIE

- [01] L. Chergui, 'Combinaison de classifieurs pour la reconnaissance de mots arabes manuscrits', Thèse de Doctorat, Université de Constantine, 2013.
- [02] R. Plamondon and G. Lorette. Automatic signature verification and writer identification – the state of the art. In *Pattern Recognition*, volume 22, pages 107–131, 1989.
- [03] R Plamondon, Neuromuscular studies of handwriting generation and representation. International Conference on Frontiers in Handwriting Recognition, 261, Kolkata, India, 2010.
- [04] Y Tang, X Wu, W Bu, Offline text-independent writer identification using stroke fragment and contour-based features. IEEE International Conference on Biometrics, 2013.
- [05] M Naghibolhosseini, F Bahrami, A behavioral model of writing. International Conference on Electrical and Computer engineering, pp: 970–973, 2008.
- [06] HN Champa, KR Anandakumar, Automated human behavior prediction through handwriting analysis. First International Conference on Integrated Intelligent Computing, pp: 160–165, 2010.
- [07] Ramírez G. et al. (2019) Overview of the Multimedia Information Processing for Personality & Social Networks Analysis Contest. In: Zhang Z., Suter D., Tian Y., Branzan Albu A., Sidère N., Jair Escalante H. (eds) *Pattern Recognition and Information Forensics. ICPR 2018. Lecture Notes in Computer Science*, vol 11188. Springer, Cham
- [08] Chawki Djeddi. : « Contribution à l'analyse et la caractérisation de l'écriture manuscrite. » PhD thesis, University of Tebessa, (June 2014).
- [09] Hemlata, Manoj Sachan, Shailendra Kumar Singh. : "Personality Detection using Handwriting Analysis", Proc. of The Seventh International Conference on Advances in Computing, Electronics and Communication - ACEC 2018
- [10] A. Schlapbach, H. Bunke. Using HMM Based Recognizers for Writer Identification and Verification, Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR-9 2004).
- [11] Kinjal Chaudhari, Ankit Thakkar. : "Survey on handwriting-based personality trait identification", *Expert Systems With Applications* 124 (2019) 282–308.
- [12] Aouraghe Ibtissame, Ammour Alae, Aboulem Ghita, Khaissidi Ghizlane, Mrabti Mostafa1, Belahsen Faouzi, Mounim A. El-Yacoubi, Sonia Garcia-Salicetti. : «Étude de l'écriture manuscrite sur tablette graphique pour l'aide au diagnostic précoce

des maladies neurodégénératives. Premiers résultats chez des patients parkinsoniens », Revue Neurologique Volume 174, Supplement 1, April 2018, Page S33.

- [13] Champa H N, K R AnandaKumar, "A Scientific Approach to Behavior Analysis through Handwriting Analysis", National Conference on Research Trends in Information Technology, S R K R Engineering College, Bhimavaram, Andhra Pradesh, 2008.
- [14] N Mogharreban, S Rahimi, M Sabharwal, "A Combined Crisp and Fuzzy Approach for Handwriting Analysis", IEEE Annual Meeting of the Fuzzy Information, 2004, vol1, pp 351-356.
- [15] Sung-Hyuk and Charles C Tappert , "Automatic Detection of Handwriting Forgery", Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, 2002, vol1, pp.351-356.
- [16] Gabriela Ramírez, Esaú Villatoro, Bogdan Ionescu, Hugo Jair Escalante, Sergio Escalera, Martha Larson, Henning Müller, and Isabelle Guyon. : " Overview of the Multimedia Information Processing for Personality & Social Networks Analysis Contest".
- [17] Costa, P.T. Jr., McCrae, R.R., Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual, Psychol. Assess. Resources, Odessa, FL, 1992.
- [18] J. Richiardi and A. Drygajlo. Gaussian mixture models for on-line signature verification. In Intl Multimedia Conf., Proc. 2003 ACM SIGMM workshop on Biometrics methods and applications, pages 115-122, Berkeley, USA, Nov. 2003.
- [19] Edson J. R. Justino, Abdenaim El Yacoubi, Fl avio Bortolozzi, and Robert Sabourin. An off-line signature verification system using hidden markov model and cross-validation. In SIBGRAPI, pages 105-112, 2000.
- [20] M. Amrouch, 'Reconnaissance de caractères imprimés et manuscrits, textes et documents basée sur les modèles de Markov cachés', Thèse de doctorat, Université Ibn Zohr Maroc, 2014.
- [21] C. Djeddi, L-S. Meslati, I. Siddiqi, A. Ennaji, H. El Abed, A. Gattal. Evaluation of Texture Features for Offline Arabic Writer Identification, Conference Paper · April 2014.
- [22] D. Bertolini, L.S. Oliveira, E. Justino, R. Sabourin. "Texture-based descriptors for writer identification and verification". Expert Systems with Applications 40 (2013) pages 2069–2080.
- [23] A. Varshney and S. Puri, "identification on the basis of Handwriting," pp. 1–6, 2017.
- [24] A. R. M and D. Varghese, "HABIT : Handwritten Analysis Based Individualistic Traits Prediction," no. 7, pp. 209–218, International Journal of Image Processing (IJIP), Volume (7): Issue (2): 2013.

- [25] P. S. Kedar, M. V. Nair, and M. S. Kulkarni, "Personality Identification through Handwriting Analysis: A Review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 1, pp. 548–556,
- [26] H. N. Champa and K. R. AnandaKumar, "Automated human behavior prediction through handwriting analysis," *Proc. - 1st Int. Conf. Integr. Intell. Comput. ICIIC 2010*, pp. 160–165, 2010.
- [27] S. Lee, "Individuality of Handwriting" pp. 106–109, 2001.
- [28] "The History of Graphology _ British Institute of Graphologists," 1983. [Online]. Available: <http://www.britishgraphology.org/about-british-institute-ofgraphologists/the-history-of-graphology/>. [Accessed: 15-Apr-2018].
- [29] D. J. Antony, "Personality Profile Through Handwriting Analysis," pp. 1–118, 2008.
- [30] S. Mukherjee, "Feature Extraction from Handwritten Documents for Personality Analysis." 2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE)
- [31] A. Dhadwal, S. Alone, and R. Agarwal, "Automatic Emotion Recognition through," pp. 811–816, ICCUBEA '15 Proceedings of the 2015 International Conference on Computing Communication Control and Automation
- [32] P. Joshi, "Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach," *International journal of Computer Applications (0975-8887) volume 130-No.15 Novembre 2015*.
- [33] S. Mutalib, R. Ramli, S. A. Rahman, M. Yusoff, and A. Mohamed, "Towards emotional control recognition through handwriting using fuzzy inference," *Proc. - Int. Symp. Inf. Technol. 2008, ITSIM*, vol. 2, no. 1997, pp. 0–4, 2008.
- [34] P. K. Grewal and D. Prashar, "Behavior Prediction Through Handwriting Analysis," *International Journal of Computer Science And Technology IJCST Vol. 3, Issue 2, April - June 2012*, pp. 13–17, 2016.
- [35] K. V, B. Ottappurakkal, S. Suresh, and S. RS, "Personality Identification Based on Handwriting," *Int. J. Emerg. Technol. Comput. Appl. Sci. (IJETCAS)*, vol. 12, no. 3, pp. 231–235, 2015.
- [36] S. H. Cha and C. C. Tappert, "Automatic detection of handwriting forgery," *Proc. - Int. Work. Front. Handwrit. Recognition, IWFHR*, pp. 264–267, 2002.
- [37] V. Kamath, N. Ramaswamy, P. N. Karanth, V. Desai, and S. M. Kulkarni, "DEVELOPMENT OF AN AUTOMATED HANDWRITING ANALYSIS SYSTEM," vol. 6, no. 9, pp. 135–140, 2011.

- [38] K. Amend and M. S. Ruiz, *Handwriting Analysis The Complete Basic Book*. 1980.
- [39] N Grace, PG Enticott, BP Johnson, NJ Rinehart, Do handwriting difficulties correlated with core symptomology, motor proficiency and attentional behaviors. *Journal of Autism and Developmental Disorders*, 1–12 (January 2017)
- [40] Guillaume JOUTEL, Véronique ÉGLIN, Hubert EMPTOZ. : « Une nouvelle approche pour indexer les documents manuscrits anciens »
- [41] José E. Valdez-Rodríguez, Hiram Calvo, Edgardo M. Felipe-Riverón. : "Handwritten texts for Personality Identification Using Convolutional Neural Networks"
- [42] <http://chalearnlap.cvc.uab.es/challenge/27/track/29/description/>
- [43] Vítor C. Vidal. : "Staffline detection and removal in the Grayscale Domain" 26 June 2012
- [44] Mihai Gavrilescu et Nicolae Vizireanu. : Predicting the Big Five personality traits from handwriting. *EURASIP Journal on Image and Video Processing* (2018) 2018:57
- [45] Zhaoxiang Zhang, David Suter, YingliTian, Alexandra BranzanAlbu, Nicolas Sidère, Hugo Jair Escalante, "2018 ICPR Multimedia Information Processing for Personality & Social Networks Analysis Challenge – HwxPI Task 2-HWxPI: Handwritten texts for Personality Identification." ,In: *Pattern Recognition and Information Forensics:ICPR 2018 International Workshop, CVAUI, IWCF, and MIPPSNA Beijing, China, August 20-24, 2018*.
- [46] M. Bulacu, L. Schomaker, L. Vuurpijl, "Writer identification using edge-based directional features", *Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003)*, IEEE Computer Society, pp. 937 – 941, vol. II, Edinburgh, Scotland, 2003.
- [47] C. Djeddi, L. Souici-Meslati, "A texture based approach for Arabic Writer Identification and Verification", *IEEE International Conference on Machine and Web Intelligence, ICMWI'2010, Algiers, Algeria*, pp: 115-120, 2010.
- [48] K. Deguchi, "Two-dimensional Auto-regressive Model for Analysis and Sythesis of Gray-level textures," in *Proc. Of the 1st Int. Sym. for Science on Form*, General Ed. S. Ishizaka, Eds. Y. Kato, R. Takaki, and J. Toriwaki, pp. 441-449, Tokyo, Japan, 1986.
- [49] Garain. U., Paquet. T., "Off-line multi-script writer identification using AR coefficients," In: *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR 2009)*. Barcelona, Spain, pp. 991 – 995, 2009.
- [50] J. S. Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. F. P. da Costa, "Staff detection with stable paths," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1134–1139, 2009.

- [51] O.C Abikoye, M.A Mabayoje, R.Ajibade. : "Offline Signature Recognition & Verification using Neural Network". Volume 35-No.2, December 2011