

## RECONNAISSANCE DES CARACTERES MANUSCRITS



Groupe 4 : Vision par ordinateur (INF 4238)

Département d'informatique , Université Yaounde 1, Cameroun

\*Superviseur : Dr Hippolyte TAPAMO [hyppolyte.tapamo@facsciences-uy1.cm](mailto:hyppolyte.tapamo@facsciences-uy1.cm)

mots clés : **detection , extraction des caractéristiques, filtres .**

### Membres du Groupe.

- KAPTCHOUANG YVAN DERICK 20U2869
- NGEUKEU MEKENGUIM ROBINSON BRADLEY 20U2628
- FOKOU AROLE 20U2994

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Historique</b>	<b>3</b>
2.1	Années 1920-1960 : Les Premiers Concepts . . . . .	3
2.2	Années 1970-1980 : Avancées Technologiques . . . . .	3
2.3	Années 1990-Présent : Maturité et Adoption à Grande Échelle . . . . .	3
<b>3</b>	<b>Difficultés rencontrées</b>	<b>4</b>
<b>4</b>	<b>Prétraitement des images:</b>	<b>5</b>
4.1	Étapes du prétraitement . . . . .	5
4.1.1	Normalisation . . . . .	5
4.1.2	Binarisation . . . . .	5
4.1.3	Suppression du bruit . . . . .	6
<b>5</b>	<b>Méthodes d'extraction des caractéristiques locales</b>	<b>6</b>
5.1	Extraction des régions stable avec MOPS (Multi-Scale Oriented Patches) . . . . .	6
5.2	Detection des coins avec l'opérateur Harris . . . . .	7
5.3	Detection des coins avec l'opérateur SIFT (Scale-Invariant Feature Transform) . . . . .	7
<b>6</b>	<b>Algorithme de machine learning pour la reconnaissance des caractères manuscrits</b>	<b>8</b>
6.1	Méthode d'apprentissage classique . . . . .	8
6.1.1	Support Vector Machines (SVM) . . . . .	8
6.2	Algorithme de DeepLearning . . . . .	9
6.2.1	Convolutional Neural Network (CNN) . . . . .	9
6.2.2	Collecte des données . . . . .	9
6.2.3	Prétraitement des données . . . . .	9
6.2.4	Architecture du modèle CNN . . . . .	9
6.2.5	Vue sommaire de l'architecture du modèle . . . . .	11
6.2.6	Apprentissage . . . . .	11
6.2.7	Evaluation et validation . . . . .	11
6.2.8	Prédiction . . . . .	11
<b>7</b>	<b>Applications</b>	<b>12</b>
7.1	La numérisation des documents . . . . .	12
7.2	Reconnaissance des plaques d'immatriculation . . . . .	12
7.3	Dans le systeme éducatif . . . . .	12
<b>8</b>	<b>Conclusion</b>	<b>12</b>

## 1. Introduction

La reconnaissance de caractères manuscrits (Handwritten Character Recognition, HCR) est un domaine d'étude et de technologie visant à convertir le texte écrit à la main en un format numérique lisible par des ordinateurs. Il combine des techniques avancées de traitement d'images, d'extraction de caractéristiques et d'apprentissage automatique pour transformer le texte manuscrit en données numériques. Les progrès dans ce domaine permettent d'automatiser de nombreuses tâches autrefois manuelles et ouvrent la voie à de nouvelles applications dans divers secteurs.

## 2. Historique

### 2.1. Années 1920-1960 : Les Premiers Concepts

- **1929 : Gustav Tauschek** : Inventeur autrichien, il a breveté une machine capable de lire des caractères imprimés en utilisant des matrices et des cellules photoélectriques. Bien que rudimentaire, cette machine est l'un des premiers dispositifs d'OCR.
- **1933 : Emanuel Goldberg** : Un ingénieur russo-allemand a inventé une machine capable de lire des caractères et de convertir des textes en signaux télégraphiques.
- **1949 : David H. Shepard** : Un inventeur américain a développé le "Gismo", un appareil capable de lire des caractères imprimés et de les convertir en code Morse. C'est l'un des premiers dispositifs d'OCR à être réellement utilisé dans des applications pratiques.
- **1959 : Kurzweil Computer Products** : Raymond Kurzweil, un pionnier de la reconnaissance des caractères, a développé une machine qui pouvait lire des textes imprimés et convertis en parole, ouvrant la voie à la reconnaissance des caractères pour les aveugles.

### 2.2. Années 1970-1980 : Avancées Technologiques

- **1974 : Kurzweil Reading Machine** : Raymond Kurzweil a développé la première machine de lecture pour les aveugles, capable de reconnaître des caractères variés et de les convertir en synthèse vocale.
- **Années 1980 : Progrès en informatique** : L'augmentation de la puissance de calcul et le développement des algorithmes de traitement d'image ont permis des avancées significatives dans la précision et la rapidité des systèmes OCR.

### 2.3. Années 1990-Présent : Maturité et Adoption à Grande Échelle

- **1990 : OCR pour les utilisateurs de PC** : Avec la popularisation des ordinateurs personnels, les logiciels OCR tels qu'OmniPage et FineReader ont rendu la technologie accessible aux consommateurs.
- **Années 2000-2010 : OCR et Internet** : La numérisation massive des livres et des documents, notamment par Google Books, a utilisé des technologies OCR avancées pour créer des bases de données de textes accessibles en ligne.
- **Technologies modernes** : Les techniques de deep learning et les réseaux de neurones convolutionnels (CNN) ont considérablement amélioré la précision et la capacité des systèmes OCR modernes, permettant la reconnaissance de textes dans des images complexes et de qualité variable.

### 3. Difficultés rencontrées

Malgré les avancées significatives dans le domaine de la reconnaissance optique de caractères (OCR) pour les textes imprimés, la reconnaissance de l'écriture manuscrite reste un défi majeur.

- **Variabilité de l'écriture** : La grande variabilité de l'écriture manuscrite entre différentes personnes pose un défi majeur.
- **Bruitage et qualité d'image** : Les images de mauvaise qualité ou bruitées peuvent affecter la précision des systèmes de reconnaissance.
- **la diversité des alphabets et la complexité des formes manuscrites** rendent cette tâche particulièrement difficile.
- **les variations individuelles, telles que les différences de pression, d'angle et de fluidité de l'écriture**, ajoutent une couche de complexité supplémentaire.

## 4. Prétraitement des images:

Le prétraitement des images est une étape cruciale dans la reconnaissance de caractères manuscrits (RCM). Il vise à améliorer la qualité des images afin de faciliter l'extraction des caractéristiques et la reconnaissance des caractères. Les principales étapes du prétraitement comprennent la normalisation, la binarisation et la suppression du bruit. Ces techniques permettent de rendre les images plus homogènes et d'éliminer les artefacts qui pourraient perturber les algorithmes de reconnaissance.

### 4.1. Étapes du prétraitement

#### 4.1.1. Normalisation

La normalisation des images consiste à ajuster les dimensions et l'échelle des images pour garantir une uniformité dans l'ensemble de données. Cela implique souvent le redimensionnement des images à une taille standard et l'ajustement des niveaux de gris ou de couleur.

- **Redimensionnement** : Cette opération consiste à redimensionner toutes les images à une taille fixe, par exemple 28x28 pixels. Cela permet de garantir que les réseaux de neurones reçoivent des entrées de dimensions constantes.
- **Ajustement des Niveaux de Gris** : Pour les images en niveaux de gris, cette étape implique la mise à l'échelle des valeurs de pixel pour qu'elles se situent dans une plage définie (par exemple, de 0 à 255). Cela peut également inclure la normalisation des valeurs de pixel pour qu'elles se situent entre 0 et 1.

#### 4.1.2. Binarisation

La binarisation est le processus de conversion d'une image en niveaux de gris en une image binaire, où chaque pixel est soit noir soit blanc. Cette étape est cruciale pour simplifier l'image et mettre en évidence les caractéristiques essentielles des caractères manuscrits.

- **Seuil Global** : Un seuil fixe est appliqué à tous les pixels de l'image. Par exemple, tous les pixels avec une intensité supérieure à un certain seuil sont définis comme blancs, et ceux en dessous sont définis comme noirs.
- **Seuil Adaptatif** : Le seuil est déterminé localement pour différentes régions de l'image, ce qui peut être utile pour les images avec des variations d'éclairage.

### 4.1.3. Suppression du bruit

Les images manuscrites contiennent souvent du bruit sous forme de taches, de bavures ou d'irrégularités. La suppression du bruit est essentielle pour améliorer la clarté des caractères.

- **Filtres de Lissage** : Les filtres de moyenne ou les filtres gaussiens peuvent être utilisés pour lisser l'image et réduire le bruit aléatoire.
- **Filtres Médians** : Les filtres médians sont efficaces pour éliminer les taches de bruit tout en préservant les contours des caractères.
- **Morphologie Mathématique** : Des opérations comme l'érosion et la dilatation peuvent être utilisées pour éliminer le bruit et affiner les contours des caractères.

## 5. Méthodes d'extraction des caractéristiques locales

### 5.1. Extraction des régions stable avec MOPS (Multi-Scale Oriented Patches)

MOPS (Multi-Scale Oriented Patches) est une méthode de description de caractéristiques utilisée en vision par ordinateur pour extraire et décrire des points d'intérêt dans des images. MOPS est conçu pour être invariant aux transformations géométriques telles que la rotation, le changement d'échelle et les petites déformations.

Elle permet de détecter les régions dans une image qui restent stables sur une large gamme de seuils de segmentation. En d'autres termes, ces régions ne changent pas beaucoup même si on modifie le seuil utilisé pour les segmenter. Cela les rend particulièrement robustes et utiles pour la détection d'objets.

#### Principe de MOPS

- **Détection des points d'intérêt** : Utilise un détecteur de points d'intérêt, comme Harris, pour identifier des points caractéristiques dans l'image.
- **Extraction de patches** : Pour chaque point d'intérêt détecté, un patch d'image centré sur ce point est extrait. Ce patch est généralement de petite taille, par exemple 40x40 pixels.
- **Normalisation** : Le patch extrait est normalisé pour être invariant aux transformations d'échelle et de rotation. Cela implique de redimensionner le patch à une taille fixe et de l'orienter de manière cohérente avec l'orientation dominante de la région.

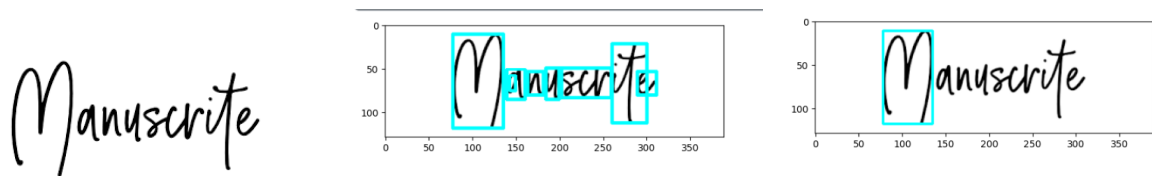


Figure 1. Extraction des régions stable avec MOPS

## 5.2. Detection des coins avec l'opérateur Harris

Le détecteur de coins de Harris est un algorithme utilisé en vision par ordinateur pour détecter les coins dans une image. Les coins sont des points d'intérêt qui présentent des changements importants dans plusieurs directions, ce qui les rend utiles pour des tâches comme la reconnaissance d'objets, la correspondance d'images et le suivi de mouvements.

Principe de fonctionnement

- Calculer le gradient en chaque point de l'image
- Créer la matrice H à partir des entrées du gradient
- Calculer les valeurs propres
- Trouver des points avec une réponse importante (  $\lambda > \text{seuil}$  )
- Choisir les points où  $\lambda$ - est un maximum local comme caractéristiques

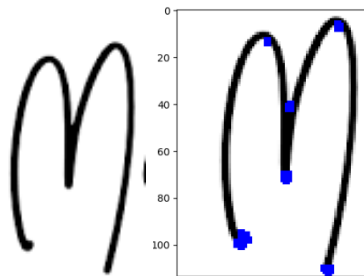


Figure 2. detection des coins avec Harris

## 5.3. Detection des coins avec l'opérateur SIFT (Scale-Invariant Feature Transform)

SIFT (Scale-Invariant Feature Transform) est une technique robuste et largement utilisée en vision par ordinateur pour la détection et la description des caractéristiques locales dans des images. L'algorithme SIFT est conçu pour être invariant aux transformations d'échelle, de rotation et partiellement invariant aux changements d'illumination et aux transformations affines

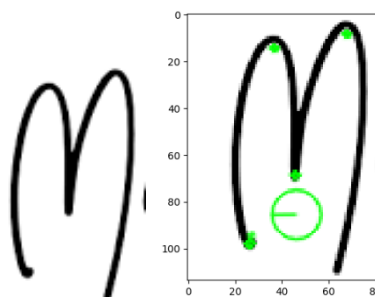


Figure 3. detection des coins avec SIFT

## 6. Algorithme de machine learning pour la reconnaissance des caractères manuscrits

### 6.1. Méthode d'apprentissage classique

#### 6.1.1. Support Vector Machines (SVM)

Il s'agit d'un algorithme de machine learning supervisé utilisé principalement pour des tâches de classification, mais aussi pour des tâches de régression.

Le SVM cherche à trouver un hyperplan dans un espace à  $n$  dimensions (où  $n$  est le nombre de caractéristiques) qui sépare les points de données appartenant à différentes classes. L'objectif est de choisir un hyperplan qui a la plus grande marge, c'est-à-dire la plus grande distance entre les points de données de chaque classe les plus proches de l'hyperplan (appelés vecteurs de support).



Figure 4. Extraction des caracteristiques avant de passer au modèle

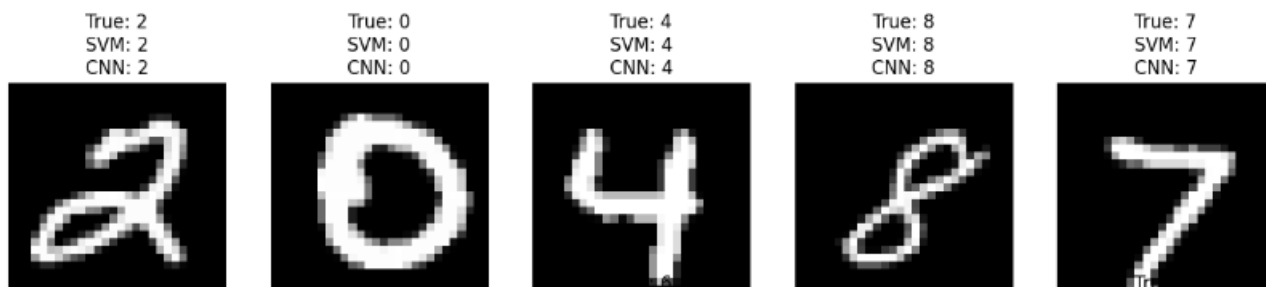


Figure 5. Prédiction sur le modèle SVM



## 6.2. Algorithme de DeepLearning

Ces algorithmes ont la capacité d'apprendre eux-mêmes les caractéristiques à extraire sur les images pour la reconnaissance des caractères manuscrit.

### 6.2.1. Convolutional Neural Network (CNN)

Ils sont utilisés pour la reconnaissance de caractères manuscrits en raison de leur capacité à extraire des caractéristiques spatiales complexes des images. Un réseau typique peut inclure plusieurs couches convolutives suivies de couches de pooling et de couches entièrement connectées.

Les étapes de mise en oeuvre sont les suivantes :

### 6.2.2. Collecte des données

Le jeu de données EMNIST contiennent des images de caractères manuscrits étiquetées avec la classe correspondante (par exemple, '0', '1', ..., '9' ; 'a'...'z' ).

### 6.2.3. Prétraitement des données

1. Redimensionnez les images (exemple : 28x28 pixels ).
2. Normalisez les valeurs des pixels (par exemple, en les divisant par 255 pour les mettre dans la plage [0, 1]).
3. Divisez les données en ensembles d'entraînement, de validation et de test.

### 6.2.4. Architecture du modèle CNN

#### 1- Les couches de convolutions

Les couches de convolution pour extraire les caractéristiques des images qu'elles reçoivent à différents niveaux.

- **Filtre convolutif:** La couche de convolution utilise un filtre convolutif, également appelé noyau ou fenêtre, qui glisse sur l'entrée. Ce filtre est généralement une matrice carrée de petits poids, définissant la relation entre les pixels voisins.
- **Convolution:** Le filtre convolutif effectue une opération de convolution sur une portion de l'entrée, multipliant chaque pixel de la portion par le poids correspondant du filtre et en additionnant les produits. Cette opération est répétée en déplaçant le filtre pixel par pixel sur toute l'entrée.
- **Fonction d'activation:** Le résultat de la convolution est ensuite passé à une fonction d'activation, comme ReLU ou la fonction sigmoïde. Cette fonction introduit une non-linéarité dans le réseau, permettant d'apprendre des relations complexes entre les entrées.
- **Sortie:** La sortie de la couche de convolution est une nouvelle image qui représente les caractéristiques spécifiques présentes dans l'entrée.

## 2- Les couches de pooling

Ces couches sont utilisées pour réduire la dimensionnalité tout en conservant les caractéristiques importantes

- **Fenetre de pooling:** La couche de pooling utilise une fenêtre de pooling carrée (par ex., 2x2, 3x3) qui se déplace sur l'entrée pixel par pixel.
- **Opération de pooling:** Dans chaque fenêtre de pooling, l'opération de pooling est appliquée. Cette opération peut prendre différentes formes, les plus courantes étant:

Max pooling: Sélectionne le pixel de valeur maximale dans la fenêtre de pooling.

Average pooling: Calcule la moyenne des valeurs de pixels dans la fenêtre de pooling.

L2-norm pooling: Calcule la norme L2 des valeurs de pixels dans la fenêtre de pooling.

- **Sortie:** La sortie de la couche de pooling est une nouvelle carte d'activation de dimension réduite, représentant les caractéristiques les plus importantes extraites par l'opération de pooling.

## 3- Couche de Flatten

Son objectif principal est de transformer une images à plusieurs dimensions qu'elle recoit en un vecteur unidimensionnel, ce qui permet de l'adapter aux couches entièrement connectées standard.

## 4- Couche entièrement connectées

Pour classifier les caractéristiques extraites.

- **Entrée:** La couche dense reçoit un vecteur unidimensionnel en tant qu'entrée. Ce vecteur est généralement le résultat d'une couche de flatten précédente, qui a transformé une "feature map" convolutive en un format compatible.
- **Multiplication matricielle:** La couche dense effectue une multiplication matricielle entre l'entrée (vecteur) et une matrice de poids. Cette matrice de poids a des dimensions (nombre de neurones de la couche dense, nombre de neurones de la couche précédente).
- **Fonction d'activation:** Le résultat final est passé à une fonction d'activation, comme ReLU ou la fonction sigmoïde. Cette fonction introduit une non-linéarité dans le réseau, permettant d'apprendre des relations complexes entre les données.
- **Sortie:** La sortie de la couche dense est un nouveau vecteur, où chaque élément correspond à l'activation d'un neurone de la couche.

### 6.2.5. Vue sommaire de l'architecture du modèle

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 26, 26, 32)	320
max_pooling2d_3 (MaxPooling2D)	(None, 13, 13, 32)	0
conv2d_4 (Conv2D)	(None, 13, 13, 64)	18,496
max_pooling2d_4 (MaxPooling2D)	(None, 6, 6, 64)	0
conv2d_5 (Conv2D)	(None, 4, 4, 128)	73,856
max_pooling2d_5 (MaxPooling2D)	(None, 2, 2, 128)	0
flatten_1 (Flatten)	(None, 512)	0
dense_3 (Dense)	(None, 64)	32,832
dense_4 (Dense)	(None, 128)	8,320
dense_5 (Dense)	(None, 26)	3,354

Total params: 137,178 (535.85 KB)

Trainable params: 137,178 (535.85 KB)

Non-trainable params: 0 (0.00 B)

### 6.2.6. Apprentissage

Parcourir les lots de données d'apprentissage:

1. Propager chaque lot d'entrée à travers le réseau, en calculant les sorties de chaque couche.
2. Calculer la perte du réseau en comparant les sorties prédites aux labels corrects.
3. Propager la perte en arrière à travers le réseau pour calculer les gradients des paramètres.
4. Mettre à jour les paramètres du réseau en suivant l'algorithme d'optimisation choisi, en utilisant les gradients calculés.

### 6.2.7. Evaluation et validation

1. Évaluer les performances du réseau sur les données de validation après chaque époque d'apprentissage.
2. Arrêter l'apprentissage si une condition de convergence est atteinte (par exemple, amélioration minimale de la performance sur la validation).

### 6.2.8. Prédiction

Utiliser le modèle entraîné pour prédire des caractères manuscrits sur des nouvelles images.

## 7. Applications

### 7.1. La numérisation des documents

Nous pouvons citer entre autres :

- **Archivage de Documents** : Transformation des documents manuscrits en fichiers numériques pour faciliter le stockage, la recherche et la gestion des archives.
- **Bibliothèques et Archives Historiques** : Numérisation des manuscrits anciens et des documents historiques pour les préserver et les rendre accessibles en ligne.

### 7.2. Reconnaissance des plaques d'immatriculation

Nous pouvons citer entre autre :

- **Contrôle du Trafic** : Utilisation de la reconnaissance de caractères pour lire les plaques d'immatriculation des véhicules dans les systèmes de gestion du trafic et de péage.
- **Sécurité et Surveillance** : Identification des véhicules dans les parkings et les zones sécurisées.

### 7.3. Dans le systeme éducatif

- **Correction Automatique des Devoirs** : Utilisation de la reconnaissance de caractères pour corriger automatiquement les devoirs et les examens manuscrits.
- **Aide à l'Écriture** : Applications qui aident les enfants à apprendre à écrire en fournissant un retour en temps réel sur leurs caractères manuscrits.

## 8. Conclusion

L'évolution rapide de la technologie et l'augmentation de la capacité de traitement des ordinateurs ont ouvert de nouvelles perspectives dans le domaine de la reconnaissance de caractères manuscrits. Ce domaine, autrefois limité par des contraintes techniques, bénéficie désormais de méthodes avancées et d'algorithmes sophistiqués qui permettent une reconnaissance plus précise et efficace des caractères manuscrits. Cette conclusion synthétise les principaux points abordés dans ce travail, en soulignant l'importance du prétraitement des images, les méthodes traditionnelles et modernes de reconnaissance de caractères, ainsi que les applications concrètes de ces techniques.