

Résumé de la mémoire de magister en électronique¹
Discrimination texte manuscrit – texte imprimé par les SVMs

ET-Tahir ZEMOURI²

Laboratoire de Communication Parlée et de Traitement du Signal (LCPTS)
Faculté d'Électronique et d'Informatique (FEI), USTHB
Spécialité : Traitement du Signal et des Images (TSI)

Résumé

Ce travail s'inscrit dans le cadre du développement d'un système de discrimination du texte imprimé et du texte manuscrit. En effet, un problème crucial dans les systèmes de reconnaissance de l'écriture est la présence de l'imprimé et le manuscrit dans la même image. Chaque type de texte doit être traité en utilisant des méthodes différentes afin d'optimiser la précision de la reconnaissance. Lorsque le document d'origine est imprimé, le mécanisme de reconnaissance est appelé OCR (Optical Character Recognition). S'il est manuscrit, la dénomination est ICR (Intelligent Character Recognition).

Notre objectif essentiel est de proposer un système de discrimination du texte manuscrit – texte imprimé afin d'appliquer des méthodes et algorithmes à chaque type d'écriture. D'une manière générale, un système de discrimination texte manuscrit - texte imprimé est composé de trois modules : Prétraitement, génération de vecteur caractéristiques et classification.

Les prétraitements appliqués sur l'image numérisée, permettent d'extraire des lignes de texte, éventuellement segmentées en mots. Cette étape comprend aussi des processus de traitement d'image, l'atténuation du bruit, la binarisation, la correction de l'inclinaison angulaire du document et le lissage de contour.

L'objectif de la génération de vecteur caractéristique est la mise en évidence de l'information pertinente qui se trouve initialement noyée dans l'information brute. Dans ce mémoire nous utilisons la transformée de Radon et les caractéristiques statistiques.

Enfin, la classification est l'élaboration d'une règle de décision qui permet d'affecter une forme à une classe. Dans notre cas, nous utilisons les SVMs pour leur robustesse comparativement aux autres classifieurs.

Le système est évalué sur la base de données IAM, les taux de reconnaissance obtenus supérieurs à 99%. Ce taux montre l'efficacité et la performance de l'approche proposée comparativement aux méthodes existantes.

Un autre avantage de ce système est qu'il est complètement automatique et il n'exige aucune intervention de l'utilisateur.

¹ Directeur de mémoire : Professeur Youcef CHIBANI (USTHB Alger).

² Ingénieur d'état en électronique, Option : Communication (Université de M'sila), 2008.