

COMBINAISON D'ALGORITHMES POUR LA RECONNAISSANCE DES CHIFFRES ET DES LETTRES BATONS DANS UN ENVIRONNEMENT MULTISCRIPTEUR D'ECRITURE COURANTE MIXTE

H. Oulhadj, J. Lemoine, E. Petit, H. Wehbi

Laboratoire d'Etude et de Recherche en Instrumentation, Signaux et Systèmes

Université Paris XII – Val de Marne

UFR de Sciences et Technologie

61, Av. du Général de Gaulle

94010 CRETEIL CEDEX – France

Tél. : 01.45.17.14.97 – Télécopie : 01.45.17.14.92 – Email : OULHADJ@univ-paris12.fr

RESUME

La reconnaissance automatique de l'écriture naturelle est une opération particulièrement complexe et ardue. A ce jour, il n'existe aucun algorithme capable de traiter de façon fiable l'intégralité des tracés qui la constituent : un texte peut contenir à la fois des chiffres, des lettres minuscules, des majuscules et des caractères bâtons. Par conséquent, il est capital de savoir extraire et séparer ces différentes entités afin de les traiter grâce aux algorithmes spécifiques à chacune des catégories d'écriture qu'ils constituent. Ce papier décrit un système en-ligne capable de réaliser cette opération grâce à la segmentation-reconnaissance des chiffres et des lettres bâtons dans un environnement multiscriteur d'écriture mixte. Pour minimiser les risques d'erreur de localisation de ces tracés, la segmentation et la reconnaissance ne sont pas séparées mais coopèrent en interagissant dans une opération unique à mesure que le scribeur écrit. Une fois les caractères alphanumériques sont segmentés et reconnus, les parties cursives restantes peuvent être facilement localisées et par suite traitées grâce aux algorithmes spécifiques à ce type de tracé.

MOTS CLES :

Écriture mixte, caractères alphanumériques, segmentation, coopération, reconnaissance en-ligne multiscriteur, combinaison d'algorithmes, classifieur.

ABSTRACT

The automatic recognition of natural handwriting is a particularly complex and arduous operation. Today, there is no algorithm able to recognize successfully all the characters that can be met in it : a text usually contains both numbers, baton characters, lowercase and capital letters. Therefore, we have to know how to extract and separate all these tracings in order to process them with specific algorithms to their categories handwriting. This paper describes an on-line system which provides a

segmentation module able to separate alphanumerical characters from lowercase letters in natural handwriting environment. In order to reduce the segmentation errors, the recognition is not separated from the segmentation but cooperate together in a single operation as the writer writes. Once alphanumerical characters are segmented and recognised, cursive parts can be easily localised and as result processed with appropriate algorithms for them.

KEYWORDS

Mixed handwriting, alphanumerical characters, segmentation, cooperation, multiscriteur on-line recognition, algorithms combination, classifier.

1 INTRODUCTION

La reconnaissance automatique de l'écriture manuscrite représente un vaste champ d'étude dont les premiers travaux datent de plusieurs dizaines d'années [4, 8, 14, 16]. Aujourd'hui, malgré les nombreuses années d'investigation consacrées au sujet, il n'existe toujours pas de système fiable capable de traiter l'écriture naturelle dans sa globalité. En effet, les résultats publiés montrent que les taux de reconnaissance obtenus sont restreints à des domaines d'application bien limités (adresses postales, chèques bancaires) ou à des catégories d'écriture très contraintes ne représentant qu'un aspect particulier de l'écriture courante et spontanée [12, 15, 23, 24]. Le travail présenté ici entre dans le cadre d'un programme d'étude visant à terme l'élimination de la majorité des contraintes imposées afin d'aborder l'écriture naturelle dans sa globalité [18, 19, 25]. Nous présentons dans cet article les parties réservées à la segmentation et à la reconnaissance en-ligne des chiffres et des lettres bâtons dans un environnement multiscriteur d'écriture mixte. Cette étape préalable, qui est non triviale par suite des difficultés sous-jacentes, s'avère tout à fait indispensable avant d'aborder la reconnaissance du mot dans sa globalité. En effet, une fois les caractères bâtons sont segmentés et

