Classification of Printed Moroccan Town and Village Names

Said Nouri, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco

Mohamed Fakir, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco

ABSTRACT

This paper presents a new method called density weight and zigzag sequence to recognize printed Arabic names. This technique was performed on two steps, the first aims to reduce matrix size of 96x96 into 12x12 using density weight techniques, in the second step the last matrix (12x12) was used to extract 144 sequences following path zigzag technique. 144 features found are used for representing each name in data set. This proposed technique was tested on Morocco town and village names using KNN with consensus rule and SVM classifiers. The perfect score was obtained with KNN (k=9) and SVM (linear kernel).

Keywords: Density Weight, KNN Consensus, Linear Kernel, Moroccan Town and Village, Printed Arabic Names, SVM, Zigzag Sequences

1. INTRODUCTION

Arabic words recognition is one of the approaches to recognize Arabic script. The global approach performs the text recognition by recognizing the whole word or sub words after segmentation of text into words or sub words (Lawgali et al., 2001; Miled et al., 1997).

Many contributions and several approaches and feature extraction techniques was proposed (Al-Hashim et al.; Alsaif et al., 2011; Menasri et al.; Abanda et al., 2009; Hachour, 2004) for Arabic words recognition that shown encouraging results than other analytic approaches because word recognition does not suffer much problems that knows other approaches such as the strong resemblance between the unit to recognize.

In this paper we propose a recognition system of printed Arabic names of Moroccan towns and villages shown in Figure 1. The proposed system is shown in Figure 2.

The proposed system consists of four steps: acquisition, reprocessing, extraction and classification step. In the first step image is scanned then in next step is converted into a binary image

DOI: 10.4018/jitr.2014100101
then resized to 96x96, unnecessary pixels which are not part of the name, which exist in outside of name area is deleted then the resulting image is resized to 96 rows and 96 columns. Density weight and zigzag sequence method is used to extract 144 features to represent each name in the data set. In the classification phase, data set was created with 16000 simples for training data and 6000 for testing. In the classification step K nearest neighbor whit consensus rule and SVM are used as a classification method.

2. PREPROCESSING

The preprocessing step aims to prepare scanned image for next steps, in this step unnecessary information was removed and the quality of meaningful information was improved.

After scanning image of the name of Moroccan town or village with adequate resolution, binarization, noise removed and localization was performed on scanned image.

2.1. Binarization

Binarization technique (Kong et al., 1996; Pratt, 1991) aims to transform the matrix of scanned image to a binary matrix. The scanned image is represented by a matrix of values vary between 0 and 255. The last matrix is divided by 255 to obtain a matrix with values between 0 and 1. The binary matrix was obtained using thresholding technique with a threshold equal 0.3.

2.2. Skew Detection and Correction

2.2.1. Skew Detection

There are several commonly used methods (Shridhar et al., 1995), (Sun & Si, 1997), (Vinciarelli et al., 2000) for detecting skew in an image. Some methods rely on detecting connected components (connected Components or CCs are roughly equivalent to sub words or characters) and finding the average angles connecting their centroids, others use projection profile analysis.

Figure 1. Simples Moroccan towns and villages names used

Figure 2. Process of classification system
Seaboard Stock Exchange’s Emerging E-Commerce Initiative
[www.igi-global.com/article/seaboard-stock-exchange-emerging-commerce/44519?camid=4v1a](www.igi-global.com/article/seaboard-stock-exchange-emerging-commerce/44519?camid=4v1a)

An Empirical Assessment of Technology Adoption as a Choice between Alternatives
[www.igi-global.com/article/empirical-assessment-technology-adoption-choice/37207?camid=4v1a](www.igi-global.com/article/empirical-assessment-technology-adoption-choice/37207?camid=4v1a)