Pattern Recognition 43 (2010) 2807–2816

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Handwritten character recognition through two-stage foreground sub-sampling

Georgios Vamvakas *, Basilis Gatos, Stavros J. Perantonis

Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", 153 10 Athens, Greece

ARTICLE INFO

Article history:
Received 10 February 2009
Received in revised form
11 February 2010
Accepted 23 February 2010

Keywords:
Handwritten character/digit recognition
Feature extraction
Two-stage classification

ABSTRACT

In this paper, we present a methodology for off-line handwritten character recognition. The proposed methodology relies on a new feature extraction technique based on recursive subdivisions of the character image so that the resulting sub-images at each iteration have balanced (approximately equal) numbers of foreground pixels, as far as this is possible. Feature extraction is followed by a two-stage classification scheme based on the level of granularity of the feature extraction method. Classes with high values in the confusion matrix are merged at a certain level and for each group of merged classes, granularity features from the level that best distinguishes them are employed. Two handwritten character databases (CEDAR and CIL) as well as two handwritten digit databases (MNIST and CEDAR) were used in order to demonstrate the effectiveness of the proposed technique. The recognition result achieved, in comparison to the ones reported in the literature, is the highest for the well-known CEDAR Character Database (94.73%) and among the best for the MNIST Database (99.03%)

1. Introduction

Optical Character Recognition (OCR) is a field of research in pattern recognition, artificial intelligence and machine vision. It refers to the mechanical or electronic translation of images of handwritten, typewritten or printed text into machine-editable text. Nowadays, the accurate recognition of machine printed characters is considered largely a solved problem. However, handwritten character recognition is comparatively difficult, as different people have different handwriting styles. So, handwritten OCR is still a subject of active research.

A widely used approach in OCR systems is to follow a two step schema: (a) represent the character as a vector of features and (b) classify the feature vector into classes [1]. Selection of a feature extraction method is important in achieving high recognition performance. A feature extraction algorithm must be robust enough so that for a variety of instances of the same symbol, similar feature sets are generated, thereby making the subsequent classification task less difficult [2]. On the other hand, Vapnik et al. [3] have suggested that powerful classification algorithms suffice even when given features are just sufficiently discriminative. The choice of classifier, however, is not an easy task since the classifier depends on many factors such as available training set, number of free parameters, etc. Classification methods based on learning from examples have been applied to character recognition mainly since the 1990s. These methods include statistical methods based on Bayes decision rule, artificial neural networks (ANNs), Kernel methods including Support Vector Machine (SVM) and multiple classifier combination [4–7]. So, taking into account all the above, we can state that feature extraction techniques, classification methods and architectures interact in complex ways.

Feature extraction methods for handwritten characters and digits have been based mainly on two types of features: (a) statistical derived from statistical distribution of points and (b) structural. The most common statistical features used for character representation are: (a) zoning, where the character is divided into several zones and features are extracted from the densities in each zone [8] or from measuring the direction of the contour of the character by computing histograms of chain codes in each zone [9], (b) projections [10] and (c) crossings, that consist of the number of transitions from foreground to background pixels along horizontal and vertical lines and distances, that rely on the calculation of the distance of the first foreground pixel detected from the upper/lower (left/right) boundaries of the image along vertical (horizontal) lines [11]. Structural features are based on topological and geometrical properties of the character while encoding some knowledge of the structure of the character or of what sort of components is made up, such as maxima and minima, reference lines, ascenders, descenders, cusps above and below a threshold, strokes and their direction between two points, horizontal curves at top or bottom, cross points, end points, branch points, etc. [12]. Many feature extraction techniques along

* Corresponding author. Tel.: +30 210 6503218; fax: +30 210 6532175.
E-mail addresses: gbam@iit.demokritos.gr (G. Vamvakas), bgat@iit.demokritos.gr (B. Gatos), sper@iit.demokritos.gr (S.J. Perantonis).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.02.018

the above lines of research have been described in the literature. For example, in Blumenstein et al. [13], a feature extraction technique that extracts direction information from the structure of the character contours and uses two neural networks based classifiers is investigated, while Camastra and Vinciarelli [14], present an OCR methodology that relies on local features derived from zoning and global ones such as the character's aspect ration followed by a recognition procedure that combines neural gas, an unsupervised version of vector quantization where no topology of a fixed dimensionality is imposed on the network, and learning vector quantization. Singh and Hewitt [15] propose a modified Hough Transform method. Character images are divided into uniform regions that are searched for vertical, horizontal and diagonal segments. The total number of such segments is fed to the classifier. Kimura et al. [16] present a feature extraction technique calculating histograms based on chain code information followed by neural and statistical classifiers. Gader et al. [17] suggest a feature extraction scheme based on the calculation of transitions from foreground to background pixels in both vertical and horizontal directions using neural networks with back-propagation for the recognition procedure. A survey on feature extraction methods can be found in [18].

There have been quite a number of successes in determination of invariant features in handwriting and a wide range of classification methods have been extensively researched. However, as mentioned in [19], most character recognition techniques use a 'one model fits all' approach, i.e. a set of features and a classification method are developed and every test pattern is subjected to the same process regardless of the constraints present in the problem domain. It is shown that approaches which employ a hierarchical treatment of patterns can have considerable advantages compared to the 'one model fits all' approaches, not only improving the recognition accuracy but also reducing the computational cost as well. In Park et al. [19], a dynamic character recognizer is presented. The recognizer begins with features extracted in a coarse resolution and focuses on smaller sub-images of the character on each recursive pass, thus working with a finer resolution of a sub-image each time, till classification meets acceptance criteria. By employing an approach called *gaze planning*, a means of expanding only some of the nodes in a tree structure similar to quad trees [20], not all of the sub-images are subjected to further subdivision but only those where it is believed that features of interest are present. So, a feature vector is extracted for each character that has more information from those sub-images that are deemed to be more important than others. The feature vector is generated by combing all features extracted in each sub-image. These features are based on histogram of gradient and moment-based projections. In [21] the character image is subdivided recursively into smaller sub-images based on the quad tree rule. The input image is then represented by fractal codes obtained at each iteration by encoding algorithm. In [22] a feature extraction technique relied on recursive subdivisions of the image for the recognition of mathematical glyphs is introduced. Each split is based on the centre of gravity of the corresponding sub-image. The initial splitting is vertical and each level of splitting then alternates between horizontal and vertical. For each rectangular region a four dimensional feature vector is extracted consisting of the vertical or horizontal component of the centroid and the three second order central moments.

Moreover, other approaches focus on measuring the similarity/dissimilarity between shapes by mapping one character onto another [23,24]. In Belongie et al. [23] the *shape context* is presented. Each shape is represented by a set of points extracted from the contour. For each shape, a descriptor is introduced, the shape context, which is the log-polar histogram of the point. Corresponding points on two similar shapes are supposed to have the same shape context thus resulting in a bipartite graph matching problem. In [24] two characters are matched by deforming the contour of one to fit the edge strengths of the other, and a dissimilarity measure is derived from the amount of deformation needed, the goodness of fit of the edges and the interior overlap between the deformed shapes.

Most classification strategies in OCR deal with a large number of classes trying to find the best discrimination among them. However, such approaches are vulnerable to classification errors when classes of similar shapes are present since they are not easily distinguished. In [25] a two-stage classification approach is presented to detect and solve possible conflicts between characters such as 'A' and 'H' or 'U' and 'V'. During the first stage, a single classifier or ensemble of classifiers detect potential conflicts. The second processing stage becomes active only when a decision on the difficult cases must be taken. A comparative study between three different two-stage hierarchical learning architectures can be found in [26].

In our work, the idea of recursive subdivisions of the character image as in [19,22] is used as a starting point. We focus on a novel feature extraction method based on different levels of granularity. At each level, features are extracted based on the point, at the intersection of the horizontal and vertical lines, which divides the character image into four sub-images that approximately consist of the same amount of foreground pixels. Even though the feature extraction method itself is quite efficient when a specific level of granularity is used, there is more to be gained in classification accuracy by exploiting the intrinsically recursive nature of the method. This is achieved by appropriately combining the results from different granularity levels using a two-stage hierarchical approach. Initially, the level at which the highest recognition rate is achieved is used to perform a preliminary discrimination, whereas the procedure is iterated once more in order to find the level at which patterns of similar shapes, confused at the first step of the classification procedure, are best distinguished. The remainder of this paper is organized as follows. In Sections 2 and 3 the proposed OCR methodology is presented while experimental results are discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. The proposed OCR methodology

The proposed OCR methodology follows a two step schema: first a feature extraction method is applied to obtain feature vectors and then a two-stage classification scheme is performed.

2.1. Preprocessing

Before employing the proposed feature extraction technique all character images must be black and white (b/w) and normalized to an $N \times N$ matrix. In case of character/digit images that are already b/w just the size normalization step is performed under the condition that the aspect ratio is preserved. On the other hand, the character/digit images that are gray scale have to pass through a binarization step beforehand. The well-known Niblack's approach [27] was used for this step.

2.2. Feature extraction

In this section a new feature extraction method for handwritten character recognition is presented. This method is based on structural features extracted directly from the character