Engineering Applications of Artificial Intelligence 45 (2015) 405–414

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Recognition of handwritten characters using local gradient feature descriptors

Olarik Surinta *,[1], Mahir F. Karaaba [1], Lambert R.B. Schomaker [1], Marco A. Wiering [1]

Institute of Artificial Intelligence and Cognitive Engineering (ALICE), University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 14 April 2015
Received in revised form
5 June 2015
Accepted 23 July 2015
Available online 10 August 2015

Keywords:
Handwritten character recognition
Feature extraction
Local gradient feature descriptor
Support vector machine
k-nearest neighbors

ABSTRACT

In this paper we propose to use local gradient feature descriptors, namely the scale invariant feature transform keypoint descriptor and the histogram of oriented gradients, for handwritten character recognition. The local gradient feature descriptors are used to extract feature vectors from the handwritten images, which are then presented to a machine learning algorithm to do the actual classification. As classifiers, the *k*-nearest neighbor and the support vector machine algorithms are used. We have evaluated these feature descriptors and classifiers on three different language scripts, namely Thai, Bangla, and Latin, consisting of both handwritten characters and digits. The results show that the local gradient feature descriptors significantly outperform directly using pixel intensities from the images. When the proposed feature descriptors are combined with the support vector machine, very high accuracies are obtained on the Thai handwritten datasets (character and digit), the Latin handwritten datasets (character and digit), and the Bangla handwritten digit dataset.

1. Introduction

Handwritten character recognition systems have several important applications, such as zip-code recognition, writer identification for e.g. forensic research, searching in historical manuscripts, and others. For such applications, the system should be able to recognize handwritten characters written on many different kinds of documents, such as contemporary or historical manuscripts. The aim is to let the system to automatically extract and recognize the characters that are embedded in the manuscript. The quality of the manuscript is one of the factors that can improve the recognition accuracy (Gupta et al., 2011). It is essential to deal with the different problems that occur in the manuscripts, such as distortions in a character image and the background noise that can appear during the scanning process. The aim of our work is to develop new algorithms that can obtain a high recognition accuracy.

Obtaining high recognition accuracies on handwritten character datasets is a challenging problem, for which many different solutions have been proposed. Although on the standard MNIST dataset extremely high accuracies have been obtained (Meier, 2011), there are many other datasets consisting of less examples and which can be considered more difficult. These datasets are challenging due to different writing styles of the same characters, different writing persons (with differences in age, gender, and education), different writing devices, and difficulties due to background noise that appears from the printer (Surinta et al., 2012).

In this paper we emphasize the importance of the recognition of complex handwritten Thai, Bangla, and Latin scripts, for which the handwritten characters and digits are highly varying due to different shapes, strokes, curls, and concavities (Mandal et al., 2011). Some samples of the handwritten characters are shown in Fig. 1. Note that the handwritten images shown in this paper are resized to the same resolution for illustration purposes. Due to the high variety, the direct use of pixel intensities may not work very well, because there is sometimes little overlap between two handwritten images displaying the same character. Therefore, in this paper we propose to use feature extraction techniques which are robust to local displacements, but still provide discriminative feature vectors as representation of the handwritten characters. The feature extraction methods that we will make use of have also been extensively used for object recognition, namely the scale invariant feature transform (SIFT) descriptor (Lowe, 2004) and the histogram of oriented gradients (HOG) (Dalal and Triggs, 2005). This paper shows that the use of these local gradient feature descriptors to extract features from handwritten characters and digits leads to a very well performing system. High recognition performances are obtained on the challenging handwritten datasets even with a simple classifier such as the *k*-nearest neighbor method, and very high recognition accuracies are obtained when using a support vector machine classifier.

* Corresponding author. Tel.: +31 50 363 9464.
E-mail addresses: o.surinta@rug.nl (O. Surinta), m.f.karaaba@rug.nl (M.F. Karaaba), l.r.b.schomaker@rug.nl (L.R.B. Schomaker), m.a.wiering@rug.nl (M.A. Wiering).
URL: http://www.ai.rug.nl/~mrolarik (O. Surinta).
[1] ALICE http://www.rug.nl/research/alice

http://dx.doi.org/10.1016/j.engappai.2015.07.017

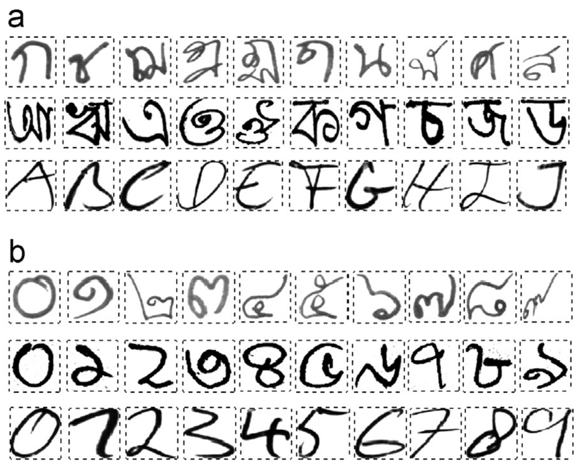

Fig. 1. Some examples of the Thai, Bangla, and Latin handwritten scripts as shown in the first, second, and third rows, respectively. Sample of (a) handwritten characters, and (b) handwritten digits.

Related work: In previous studies, the raw image (IMG) method, which directly copies the intensities of the pixels of the ink trace (Surinta et al., 2013), has often been used as the feature extraction method. It extracts a high dimensional feature vector that depends on the size of the input image.

In recent years, deep learning architectures (Hinton et al., 2006; Schmidhuber, 2015) have been effectively used for handwritten digit recognition. Most of the studies have focused on the benchmark MNIST dataset (LeCun and Cortes, 1998) and achieved high accuracies such as higher than 98% or 99%. The MNIST dataset consists of isolated handwritten digits with size of 28×28 pixel resolution and contains 60,000 training images and 10,000 test images. In Hinton et al. (2006), a greedy training algorithm is proposed for constructing a multilayer network architecture which relies on the restricted Boltzmann machine, called deep belief networks (DBN). The performance obtained from the DBN with three hidden layers (500–500–2000 hidden units) on the MNIST dataset was 98.75%. This accuracy is higher than obtained with a multi-layer perceptron and a support vector machine (SVM).

Furthermore, the convolutional neural network (CNN) (LeCun et al., 1998) is used as a feature extraction and classification technique, and the accuracy obtained is 99.47% (Jarrett et al., 2009). In another CNN-based method (Cireşan et al., 2011), the committee technique is proposed. Here multiple CNNs are combined in an ensemble, for which different CNNs are trained on different pixel resolutions of the images. The images in the dataset are rescaled from 28×28 ($N \times N$) to $N = 10, 12, 14, 16, 18$, and 20 pixel resolutions. Then, 7-net committees are used. This method obtained the high accuracy of 99.73% on MNIST. However, a single CNN in their work is reported to take approximately 1–6 h for training on a graphics processing unit (GPU) and the 7-net committees are seven times slower than a single CNN. The best technique for the MNIST dataset uses an ensemble of 35-net committees (Cireşan et al., 2012). This technique obtained the very high accuracy of 99.77%. Although such high recognition performances are sometimes achieved, these methods require large training sets and long training times to make the recognition system work well.

For handwritten Bangla digit recognition, Liu and Suen (2009) proposed to use the local gradient directions of local strokes, called the gradient direction histogram feature. The feature vectors are extracted from an image and then given to a classifier. The recognition performance of the best classifier is 99.40% on the ISI Bangla numerals dataset (Chaudhuri, 2006) which contains 19,329 training images and 4000 test images. Compared to the MNIST dataset, ISI Bangla numerals dataset is more difficult due to background noise and more different types of handwriting.

In our research, we are interested in novel methods that obtain high recognition accuracies without the availability of many training examples, and which also do not require a huge amount of training time or high performance computing algorithms.

Contributions of our paper: This paper first of all provides a new standard Thai handwritten character dataset for comparison of feature extraction techniques and methods. In this paper we will make use of three complex datasets in total, namely Bangla, Thai, and Latin, for which very high recognition accuracies have not been obtained before. This is due to the difficult problems of the Thai and the Bangla handwritten scripts such as the complex structural characteristics of the characters, the similarities between the character sets (see Fig. 7(a) and (b)), the similar structures between different characters (see Fig. 4), and the background noise. These factors negatively affect the performance of a handwritten character recognition system.

To address the problems mentioned, two local gradient feature descriptors that extract feature vectors from the challenging handwritten character images are proposed, namely the scale invariant feature transform keypoint descriptor (Lowe, 2004) and the histogram of oriented gradients (Dalal and Triggs, 2005). The feature descriptors compute feature vectors with image filters such as the Sobel filter and the Gaussian filter. Subsequently, the orientations within each region are calculated and weighted into an orientation histogram. Because these feature descriptors are invariant to small local displacements, the descriptors provide robust feature vectors.

These feature extraction methods extract features, which are then used as input for a classifier. In this paper, we experimented with two different classifiers: a k-nearest neighbor classifier and a support vector machine, so that we can also compare performance differences between these machine learning methods. We evaluate the methods on the three handwritten character scripts: Thai, Bangla, and Latin, for which we use both the handwritten characters and the handwritten digits. To show the importance of using the proposed local gradient feature descriptors, we have compared them to a method that directly uses pixel intensities of the handwritten images (called the IMG method). The results show that the feature descriptors with the support vector machine obtain very high recognition performances on the datasets, whereas the use of the IMG method performs much worse.

Paper outline: This paper is organized in the following way. Section 2 describes the local gradient feature descriptors. Section 3 describes the two classifiers including the k-nearest neighbors algorithm as a simple classifier and the support vector machine algorithm with the radial basis function kernel as a more powerful classifier. The handwritten character datasets which are used in the experiments, namely Thai, Bangla, and Latin scripts, are described in Section 4. The experimental results of the different combinations of feature descriptors and classifiers are presented in Section 5. The conclusion and some directions for future work are given in the last section.

2. Local gradient feature descriptors

To study the effectiveness of local gradient feature descriptors for handwritten character recognition, we compare two existing feature extraction techniques, namely the histogram of oriented gradients and the scale invariant feature transform keypoint descriptor. Moreover, these local gradient feature descriptors are compared to the IMG method. The IMG method uses the raw pixel intensities of the handwritten images and is a simple and widely used method. In this study, the handwritten images are resized to two pixel resolutions, 28×28 and 36×36, so that for the IMG method 784 and 1296 feature values are computed, respectively.