2009 10th International Conference on Document Analysis and Recognition

Isolated Handwritten Farsi numerals Recognition Using Sparse And Over-Complete Representations

W.M. Pan, T.D. Bui, and C.Y. Suen
Center for Pattern Recognition and Machine Intelligence, Concordia University
{wumo_pan, bui, suen}@cenparmi.concordia.ca

Abstract

A new isolated handwritten Farsi numeral recognition algorithm is proposed in this paper, which exploits the sparse and over-complete structure from the handwritten Farsi numeral data. In this research, the sparse structure is represented as an over-complete dictionary, which is learned by the K-SVD algorithm. These atoms in this dictionary are adopted to initialize the first layer of the Convolutional Neural Network (CNN), the latter is then trained to do the classification task. Data distortion techniques are also applied to promote the generalization capability of the trained classifier. Experiments have shown that good results have been achieved in CENPARMI handwritten Farsi numeral database.

1. Introduction

Recognition of handwritten Latin numerals has been extensively studied in the past few decades. Many methods have been proposed with very high recognition rate [20, 7, 10, 12, 22, 15, 16, 9, 13, 5]. Some of these techniques have found application in real systems, such as mail resorting [19] and bank check recognition [4, 8].

On the other hand, research progress has been very limited towards automatic recognition of numerals written in scripts other than Latin. In this paper, the recognition of handwritten numerals in one important cursive script, Farsi (Persian), has been investigated. Farsi is the main language used in Iran and Afghanistan, and it is spoken by more than 110 million people, including some people in Tajikistan, and Pakistan. Due to its wide usage, the problem of automatic recognition of handwritten Farsi script has attracted increasing interest in the research community.

Similar to Latin script, handwritten Farsi numerals have large variations in writing styles, sizes, and orientations. What makes the recognition of the handwritten Farsi numerals more challenging is that some of the numerals can be legally written in different shapes. Fig. 1 demonstrates the situation, where numeral '2', '3', '4' and '6' are written in two different shapes, respectively. Several recognition techniques for handwritten Farsi numerals have already been published. Soltanzadeh et al. [18] use outer profiles, crossing counts and projection histograms from multiple orientations as features. A Support Vector Machine is trained using these features for classification. Mowlaei et al.[14] propose a system for recognition of isolated Farsi numerals and characters. They use Haar wavelet to obtain the features and then feed these features to Support Vector Machines for training. Ziaratban et al.[23] propose a template-based feature extraction method. Twenty templates that are believed to be able to capture the most significant information from handwritten Farsi/Arabic numerals are selected heuristically. Feature extraction is carried out via template matching: for each template, find the best match in an input image and record the location and the match score of the best match as features. These features are then fed into a multi-layer perceptron for training. Liu et al. [2] have recently provided a new benchmark on the recognition of handwritten Bangla and Farsi numeral characters. In this benchmark, the best result on the CENPARMI handwritten Farsi numeral database has been achieved by the class-specific feature polynomial classifier(CFPC) [12] using 8-direction gradient features extracted from grayscale images after moment normalization. The achieved low error rate is 0.84%.

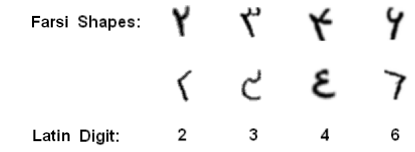

Figure 1. Numerals that can be legally written in different shapes in Farsi.

In this paper, a new handwritten Farsi numerals recogni-

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.80

tion method has been proposed. Our method tries to exploit the over-complete and sparse structure of the data. This structure is represented as an over-complete dictionary, whose atoms give sparse representations to the data. The atoms in the learned dictionary are then adopted to initialize the weights of the first layer of the CNN. The rationale behind this method is that: as will be shown later in this paper, the atoms in the learned dictionary appear to be similar to the receptive fields (maps that describe the response region of the neurons) of the visual neurons in mammalian primary visual cortex. That is, they are local bandpass filters with orientation selectivity. Initializing the first layer of the CNN with these atoms will force those artificial neurons to behave in a similar way as those neurons in the mammalian visual system.

The remainder of this paper is organized like the followings. In section 2, the CNN structure used in this research is briefly described. Section 3 introduces the K-SVD algorithm and show how it is used to learn the sparse and over-complete representations from the Farsi handwritten numerals samples. In section 4, the experimental results and the comparison between the proposed method and several commonly applied classification techniques are presented. Discussions are also presented in this section. Finally, section 5 concludes this paper.

2 Convolutional Neural Network (CNN)

A convolutional neural network is a neural network with a "deep" supervised learning architecture that has been shown to perform well in visual tasks [9, 16]. A CNN can be divided into two parts: automatic feature extractor and a trainable classifier. The automatic feature extractor contains feature map layers that extract discriminating features from the input patterns via two operations: linear filtering and down-sampling. The size of the filter kernels in the feature maps is set to 5 by 5 and the down-sampling ratio is set to 2. A back-propagation algorithm is used to train the classifier and learn the weights in the filter kernels.

Instead of using the CNN with more complicated architecture like LeNet-5 [9], a simplified implementation proposed in [16] has been chosen. The network architecture is shown in Fig. 2. The input layer is a 35 by 35 matrix containing a normalized pattern. The normalization procedure will be explained in detail in the experimental results section. The second (with N_1 feature maps) and the third (with N_2 feature maps) layers are two feature map layers, which are trained to do feature extraction at two different resolutions. Each neuron in these two layers is connected to a 5 by 5 window in the previous layer, with the strength of the connection defined by the weights in the filter kernel. Neurons in one feature map share the same filter kernel. The remaining part of the architecture consists of a fully connected multi-layer perceptron with N_3 neurons in the hidden layer and 10 neurons in the output layer. Here, we set $N_1 = 50$, $N_2 = 50$, and $N_3 = 150$.

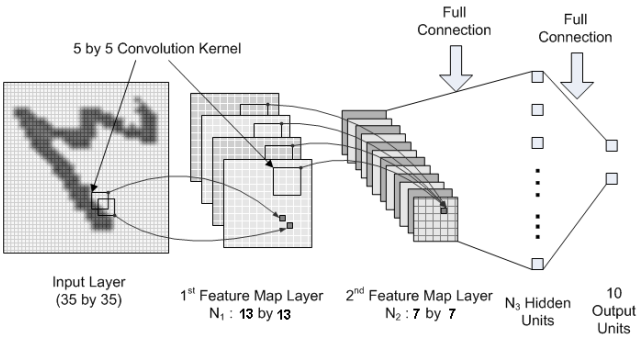

Figure 2. Adopted CNN architecture.

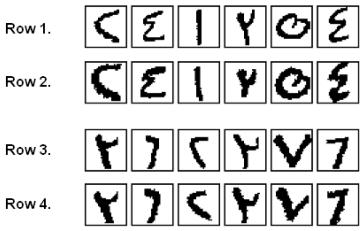

Figure 3. Some distorted samples. Rows 1 and 3: samples after preprocessing. Rows 2 and 4: the same samples after distortion.

Since the CNN architecture has a large number of weights to be learned, the number of training samples needed to train this network is also very large. In order to get better performance in terms of the capability of generalization, one usually adopted practice is to expand the training data set by introducing some distortion or transformations to the data at hand, such as affine transformations[9] and elastic distortions[16].

In this experiment, Simard's elastic distortions is combined with scaling and rotation transforms [16]. The scaling factor is selected uniformly from [-0.15, 0.15], with the negative scaling factor standing for shrinkage, while the positive scaling factor stands for enlargement. The rotation angle is also picked uniformly from $[-5°, 5°]$, with the negative angle standing for counterclockwise rotation and the positive angle standing for clockwise rotation. Figure 3 shows some examples of the distorted samples.

3 Learning Sparse And Over-Complete Representations

Suppose we have some data represented by an $n \times N$ matrix Y, where each column of Y stands for one sample. To find an over-complete dictionary D ($D \in \mathbf{R}^{n \times K}$ with $K > n$) so that

$$Y = DX \tag{1}$$

and each column of X is as sparse as possible. Each column of D is also called an *atom*. This problem can be solved using the following K-SVD algorithm.

3.1 the K-SVD Algorithm

In [1], Aharon et al. propose to solve the over-complete dictionary searching problem by working on the optimization problem below:

$$\min_{D,X}\{\|Y - DX\|_F^2\} \text{ subject to } \forall i, \|x_i\|_0 \leq T_0, \tag{2}$$

where $\|x\|_0$ gives the number of non-zero components in the vector x, x_i is the i-th column of the matrix X and T_0 is a threshold specifying the maximum number of non-zero coefficients needed for the representation. It is enforced to make sure that the learned dictionary D would give sparse representation to the data in Y. The notation $\|\cdot\|_F$ stands for the Frobenius norm.

The problem (2) is solved iteratively. First, the dictionary D is assumed to be fixed and the algorithm tries to find sparse coefficients X. Since the penalty term can be rewritten as

$$\|Y - DX\|_F^2 = \sum_{i=1}^{N} \|y_i - Dx_i\|_2^2, \tag{3}$$

problem (2) can be decoupled into N distinct problems of the form

$$\min_{x_i}\{\|y_i - Dx_i\|_2^2\} \text{ subject to } \|x_i\|_0 \leq T_0, \text{for } i = 1, 2, ..., N. \tag{4}$$

These problems are then solved using orthogonal matching pursuit (OMP) method [21].

The second part of K-SVD algorithm is to update the dictionary D. This update process is performed column by column, using singular value decomposition (SVD). The iteration procedure continues until the algorithm converges.

3.2 Learning with K-SVD

To learn the over-complete dictionary for the Farsi handwritten numerals, we randomly selected 26,156 of 5×5 patches from the sample database after preprocessing. The size of the image patch is chosen to be the same as the filter kernel in the first feature map layer of the CNN. We have chosen $K = 50$, which means that the dictionary D has a redundancy factor of 2. Another parameter to be specified is T_0, which is chosen experimentally to be 7. The atoms in the learned over-complete dictionary are shown in Fig. 4. These atoms in the learned dictionary possess significant orientation and scale selectivity, which are desired properties for feature extraction.

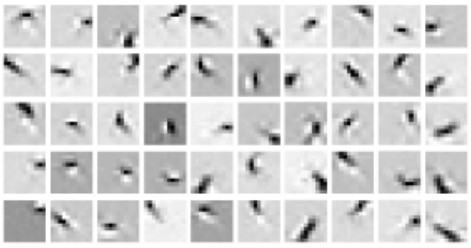

Figure 4. Over-complete dictionary learned with $T_0 = 7$.

4 Experiments

To evaluate its effectiveness, the proposed method has been applied on the CENPARMI Handwritten Farsi numeral database [17] (CENPARMI database in short). We briefly describe the properties of this database and show some of its samples. Then, the preprocessing procedures that have been applied to these data is explained. Finally, comparison of the proposed method with two different types of classifiers, the Support Vector Machines (SVM) [3] and the Modified Quadratic Discriminant Function (MQDF) [6], is made and the results are presented.

4.1 The CENPARMI Database

The Farsi handwritten numeral database used in our experiments has been built at CENPARMI [17]. These samples are collected from 175 writers of different ages, education and genders. All these samples are scanned as 300dpi color images and then converted into grayscale. These samples are further divided into non-overlapping training, verification and testing sets. There are 11,000 samples in the training set, with 1,100 samples for each class; 2,000 samples in the verification set, with 200 samples for each class and 5,000 samples in the testing set, with 500 samples for each class. Some examples of the samples in this database are shown in Fig. 5.

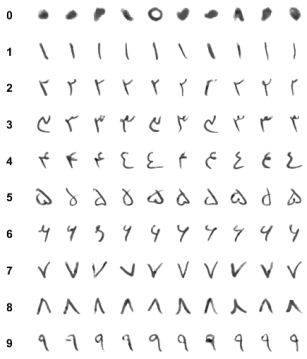

Figure 5. Examples of the samples in the CENPARMI database.

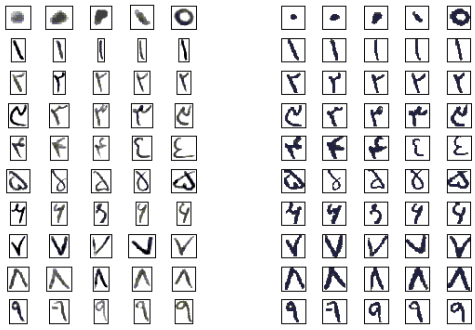

Figure 6. Examples of the preprocessing results. Left: original samples; Right: samples after preprocessing.

4.2 Preprocessing

As shown in Fig. 5, there are large variations in both the grayscale values and the sizes of the samples in the CENPARMI database. Therefore, grayscale normalization and size normalization techniques are needed.

For grayscale normalization, the gray levels of the foreground pixels of each input image have been re-scaled so that the scaled values give a standard mean of 210 and deviation of 20.

As to the size normalization, the moment normalization technique [11] is used, which first aligns the centroid of a character image with the geometric center of the normalize plan, and then re-frames the character using second-order moments. This method works better than traditional linear normalization technique in handwritten character recognition, since it is able to reduce the position variation of the important feature points in the image. Furthermore, it can cut the tails of some elongated strokes in the pattern and thus retain most classification related information. Some examples of the numerals before preprocessing and after preprocessing are given in Fig 6. Each pattern is normalized to size 35 by 35.

Table 1. Test error rates (%) of the classifiers investigated in this paper.

	MQDF	SVM	Proposed Method
Gradient Features	2.12	1.02	0.78
Profile Features	3.18	2.68	

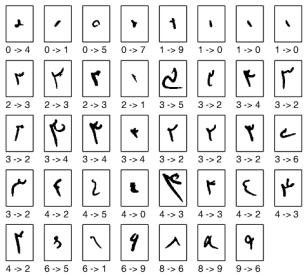

Figure 7. Misrecognized samples by the proposed method.

4.3 Experimental Results

In the experiment, the CNN has been trained with its first feature map layer initialized with the atoms in the learned over-complete dictionary. The training procedure stops after 90 epochs.

As comparison, other two well-known classifiers have also applied to the same database: SVM and MQDF. Different from CNN, these two methods require extracted features as input. Two sets of features have been investigated in this research: *gradient features* [11] and *profile features* [18]. Results of these experiments are shown in Table 1. The misrecognized samples of the proposed method are shown in Fig. 7.

4.4 Discussion

From the error cases shown in Fig. 7, we can identify the following common error cases: a) '2' - '3' misclassification, b) '3' - '4' misclassification and c) '1' - '0' misclassifications. Most of these misclassified samples are very similar in shape. Some are due to the imperfect samples in the data,

which might be introduced by the writers participating the data collection process.

The error rates in Table 1 show that, generally speaking, MQDF does not perform as well as SVM or the proposed method. This is because MQDF usually works well for data whose underlying distribution is Gaussian. However, the deviation from Gaussian distribution of the handwritten Farsi numerals are significant, especially for digits like '2', '3', '4' and '6', where each pattern has more than one representative shapes. The proposed method gets the best results, where the CNN has been initialized with the learned over-complete representations.

5 Conclusion

In this paper, a new handwritten Farsi numeral recognition method has been proposed that makes use of the sparse and over-complete structure within the data. The proposed method has been applied to a publicly available handwritten Farsi numeral database: the CENPARMI Farsi numeral database. Comparison between the proposed method and two other popular classifiers (SVM and MQDF) has also been made. These experimental results have justified the benefit of exploiting the sparse and over-complete structure in the data in the pattern recognition problems.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006.

[2] C.-L.Liu and C.Y.Suen. A new benchmark on the recognition of handwritten bangla and farsi numeral characters. *Pattern Recognition, In press*, 2008.

[3] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. Software available at *http://www.csie.ntu.edu.tw/ cjlin/libsvm*, 2001.

[4] G. Dimauro, S. Impedovo, G. Pirlo, and A. Salzo. Automatic bankcheck processing: a new engineered system. *Machine Perception and Artificial Intelligence*, 28:5–42, 1997.

[5] D. Keysers, T. Deselaers, C. Gollan, and H. Ney. Deformation models for image recognition. *IEEE Trans. Patt. Anal. Mach. Intell.*, 29(8):1422–1435, 2007.

[6] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake. Modified quadratic discriminant functions and the application to Chinese character recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(1):149–153, 1987.

[7] F. Lauer, C. Y. Suen, and G. Bloch. A trainable feature extractor for handwritten digit recognition. *Pattern Recognition*, 40(6):1816–1824, 2007.

[8] Y. LeCun, L. Bottou, and Y. Bengio. Reading checks with graph transformer networks. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 151–154, 1997.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 86(11):2278–2324, November 1998.

[10] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa. Handwritten digit recognition: Benchmarking of the state-of-the-art techniques. *Pattern Recognition*, 36(10):2271–2285, 2003.

[11] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa. Handwritten digit recognition: Investigation of normalization and feature extraction techniques. *Pattern Recognition*, 37(2):265–279, 2004.

[12] C.-L. Liu and H. Sako. Class-specific feature polynomial classifier for pattern classification and its application to handwritten numeral recognition. *Pattern Recognition*, 39(4):669–681, 2006.

[13] R. Marc'Aurelio, C. Poultney, S. Chopra, and Y. LeCun. Efficient learning of sparse representations with an energy-based model. In M. Press, editor, *Proc. Advances in Neural Information Processing Systems*, 2006.

[14] A. Mowlaei and K. Faez. Recognition of isolated handwritten Persian/Arabic characters and numerals using support vector machines. In *Proc. IEEE 13th Workshop on Neural Networks for Signal Processing*, pages 547–554, 2003.

[15] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura. Handwritten numeral recognition using gradient and curvature of gray scale image. *Pattern Recognition*, 35(10):2051–2059, 2002.

[16] P. Y. Simard, D. Steinkraus, and J. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *Proc. International Conference on Document Analysis and Recognition (ICDAR)*, pages 958–962, 2003.

[17] F. Solimanpour, J. Sadri, and C. Suen. Standard databases for recognition of handwritten digits, numerical strings, legal amounts, letters and dates in Farsi language. In *Proc. 10th International Workshop on Frontiers in Handwriting Recognition*, pages 3–7, 2006.

[18] H. Soltanzadeh and M. Rahmati. Recognition of Persian handwritten digits using image profiles of multiple orientations. *Pattern Recognition Lett.*, 25(14):1569–1576, 2004.

[19] S. Srihari and E. Keubert. Integration of handwritten address interpretation technology into the United States Postal Service Remote Computer Reader system. In *Proc. Fourth International Conference on Document Analysis and Recognition*, volume 2, pages 892–896, 1997.

[20] C. Suen, K. Liu, and N. Strathy. Sorting and recognizing cheques and financial documents. In *Proc. of third IAPR workshop on document analysis systems*, pages 1–18, 1998.

[21] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.

[22] P. Zhang, T. Bui, and C. Suen. A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. *Pattern Recognition*, 40(12):3415–3429, 2007.

[23] M. Ziaratban, K. Faez, and F. Faradji. Language-based feature extraction using template-matching in Farsi/Arabic handwritten numeral recognition. In *Proc. Ninth International Conference on Document Analysis and Recognition*, volume 1, pages 297–301, 2007.