Decision Templates with Gradient based Features for Farsi Handwritten Word Recognition

Reza Ebrahimpour¹, Roya Davoudi Vahid², Babak Mazloom Nezhad²

¹ Department of Electrical Engineering, Shahid Rajaee Teachers Training University, Tehran, Iran
² Department of Electrical Engineering, Shahid Beheshti University, Tehran, Iran
ebrahimpour@ipm.ir, r.davoudi@mail.sbu.ac.ir, b-mazloom@sbu.ac.ir

Abstract

This paper proposes a new classification method for Farsi handwritten word recognition using a scale invariant gradient based features. The extracted feature vectors classified using MLPs. Outputs of MLPs are then combined by Decision Templates. The experiments were achieved using the Iranshahr dataset. This dataset consist of 780 samples of 30 city names of Iran which 600 samples for train and 180 samples for test was used. A set of experiments were conducted to compare Decision Templates with some combination rules. Results show that template based fusion method is superior to the other schemes.

Keywords: Farsi handwritten word recognition, Feature extraction, Classifier fusion, Decision Templates.

1. Introduction

Handwritten recognition can be seen as a sub-task of more general Optical Character Recognition (OCR). Providing simple interface between man and machine is advantage of handwritten recognition systems. First, recognition of isolated handwritten digits and characters was investigated [1,2]. Later recognition of whole words [3,4] was addressed. There are many published works that deal with handwritten recognition of English [5-7], Japanese [8], Chinese [9] and Arabic scripts [10,11]. Farsi/Arabic handwritten recognition has slowly proceeded due to the special characteristics of these languages. Three important steps in handwritten recognition are pre-processing, feature extraction and classification. The aim of preprocessing is to improve the quality of the images for further processing and analysis task [12]. Handwritten recognition performance largely depends on the feature extraction approach. For extract features from word’s image various approaches are proposed. The most commonly used features are: zoning [13,14], Gradient [15], Projection histogram [16], Wavelet [17], crossing points and contours [18]. Some of these methods extract high dimension feature vectors. The selection of high performing and scale invariant feature extraction method is an important but difficult task in developing Farsi handwritten word recognition (FHWR) systems. In this paper, a novel method for feature extraction presents. For extracting the features, images should be passed from the thinning stage that caused reduction of samples thickness into a single pixel [19]. Four of the 3×3 masks are applied to the thinned word images to extract horizontal, vertical, right and left-diagonal lines. Images store separately correspond to each line. Afterward, decomposed images should be partitioned to eight sectors around the center of image and the number of black pixels in each sector calculated and normalized by dividing them upon the total number of black pixels.
in word images for feature vector. This method solves the problem of scale invariance, has a low feature dimension and high recognition rate.

In recent years, many new classifiers have been proposed and tested on various OCR databases. Improving recognition performance in difficult classification problems [20,21] is objective of multiple classifiers combination. There are two main strategies in combining classifiers: classifier selection and classifier fusion [22]. Classifier selection attempts to choose the best classifier for a given task. This method assumes that a classifier is an expert on a subset of the feature space. In classifier fusion assume that all classifiers are trained over the whole feature space [23-25]. In this paper, a templates based classifier fusion method is proposed to work out the handwritten word recognition problem. First, for training samples the decision profile matrix is constructed. Decision templates are composed with averaging observed decision profiles for each class among training. The higher similarity between decision profiles of test samples and the decision template for each class is causing the higher support for that class [26]. The decision templates fusion uses all classifier outputs to calculate the final support for each class. This method is in sharp contrast to most other fusion methods which use only the support for that particular class to make their decision.

The paper is organized as follows. Section 2 introduces new feature extraction method. Section 3 describes proposed model for fusion in details. Simulations and results are discussed in section 4 and conclusions are presented in Section 5.

2. Feature Extraction

Feature extraction is a necessary and important step in achieving good performance of handwritten recognition. There are many different ways to extract and represent features from word images. In this section, some of commonly used methods are briefly described. In addition, a novel method is introduced that gives high recognition rate compared with other features.

2.1. Zoning

In this method the word image is divided into number of zones. In each zone the average of black pixels is calculated and used as feature vector [27,28].

2.2. Projection histogram

The horizontal and vertical projection histograms of the image are obtained as follows (Trier et al., 1996): for each row and column between the first and last non-empty rows and columns of the image, the number of black pixels is counted and these values are placed in separated vectors. By concatenation of horizontal and vertical projection histogram the final feature vector is achieved [29].

2.3. Gradient feature

The Sobel operators are used to generate the gradient amplitude and phases. The gradient magnitude of each pixel is quantized into eight layers corresponding to one of the Freeman directions. The image is divided into 8 sub-images and for each sub-image the number of black pixels is computed as feature [30].

2.4. Wavelet transform
Wavelet transform is a series expansion technique that allows us to represent the signal at different levels of resolution. The wavelet transformation uses the so-called wavelet basis functions (shortly wavelets) and the scaling functions, both forming the orthogonal or biorthogonal family of basis function. The wavelet function has good localization abilities in both time and frequency [31-33].

2.5. Vertical and horizontal cross count

The vertical and horizontal cross counts are calculated by scanning each column and row of the binary image respectively. Each transition from 0 to 1 or from 1 to 0 increases a counter that has an initial value of zero [29].

2.6. Proposed method

This paper proposes a new scale invariant gradient based method for Farsi Handwritten Word Recognition feature extraction. Here this method explained in details. It should be noted that for this method first, thinning must be applied on the word images. Thinning is the process of reducing thickness of samples to just a single pixel. By thinning method shape information of patterns is preserved and data size is reduced [34]. Thinning method removes pixels so that a pattern without holes shrinks to a minimally connected stroke, and a pattern with holes shrinks to a connected ring halfway between each hole and the outer boundary. In this research work, for better symbol representation morphology based thinning algorithm is used. Second, the word images decomposed into a number of separate images corresponding to four of the 3×3 masks. Indeed this masks used for scanning horizontal, vertical, right and left diagonal lines to produce their equivalent images. Fig.1 shows these masks. For example the first mask used for separating the horizontal lines from the word images. These lines can be extracted as explained below. Instead of input word image considered a zero matrix with the size of original image. This mask moved from left to right over the image and specified new values of elements of this matrix. The assigned value of this matrix depends on the value of pixels in original word image. That is, at position of (i,j), the value of (i,j), (i,j+1) and (i,j-1) pixels from matrix are the same and equal to one if all of these three pixels from sightly image have been one and zero, otherwise.

This procedure repeated for other three masks and each word image decomposed into four separate images corresponding to these masks. In the other word, with these masks horizontal, vertical, right and left diagonal lines in word images are separated. Finally, each of separated images is uniformly partitioned into 8 sectors around the image center. The number of black pixels is calculated in each sector. These values normalized by dividing them upon the total number of black pixels in word images and used for feature value of that sector. Hence for each of these separated images we have eight feature values leading to a

![Figure 1. The masks that used for decomposing word images into a number of separate images.](image-url)
feature vector of 32 elements. With this method the problem of scale invariance is solved and extracted feature vectors have low dimension. Fig.2 presents the stages of this method.

3. Classifier fusion

Classifier fusion is combination of the decision of different individual classifiers when classifying new patterns. Classifier fusion (CF) assumes that all individual classifiers are competitive, instead of complementary [35]. The reasons for combination of classifiers are the possibility to boost the classification accuracy, combining classifiers with different errors [36], or combining local experts. The classifiers should be independent or diverse in error distribution. An overview of methods for classifiers diversifying is presented in [37,38]. There are two main structures for multiple classifier fusion: static and dynamic. Static structures only look at the output of the classifiers and are quite simple. Maximum, minimum, median, average and product belong to static categories. Dynamic structures take into account information from the training phase about the behavior of the classifiers and are more elaborate [39]. Many fusion methods exist that aim to achieve this. One of them is decision templates that described in this paper.

3.1. Decision templates

The method here described is based on a set of c matrices called decision templates (DTs). For each class, Decision Templates indicates the classifier outputs for patterns which are belong to that class. For constituting these matrices all classifier outputs are used. Train samples with specified class label feed to classifiers and by combining their outputs, decision profile matrices are constructed. Decision profile (DP) is a matrix consisting of one row for each classifier with the classifier output in the rows [39]. When DPs are generated, decision templates for each class construct with the averaging of the decision profiles of the elements of the training set which pertain to that class [40]. Mathematical description of this, is:

\[
DT_c = \begin{bmatrix}
  dt_c(y,1) & \cdots & dt_c(y,M) \\
  \vdots & \ddots & \vdots \\
  dt_c(L,1) & \cdots & dt_c(L,M)
\end{bmatrix},
\]

\[
dt_c(y,z) = \frac{\sum_{i=1}^{n} Ind_c(x_i)d_{y,z}(x_i)}{\sum_{i=1}^{n} Ind_c(x_i)}
\]

Where \( d_{y,z}(x_i) \) is the degree of support given by the \( y \)-th classifier for the sample \( x_i \) of the class \( z \). \( Ind_c(x_i) \) has a value of one if the class of \( x_i \) is \( c \), and zero, otherwise [41].

In the test phase the DTs of each class is compared to the decision profile of new sample with various measures of similarity and the class with most similarity is selected. For matching between decision profile of test samples and DTs, various fuzzy similarity measures can be used. One of the most practical similarity measures is normalized Euclidean distance but any distance criteria such as the Minkowski and the Mahalanobis can be used.
4. Experimental results

The proposed feature extraction and DTs method that investigated in this research is tested on a dataset of city names of Iran(Iranshahr dataset). The Iranshahr dataset consist of 780 samples of 30 city names of Iran. For each city name 26 samples is accessible which written by 26 different persons. In this paper 20 samples had been selected for train and 6 for test, so training set consist of 600 samples and test set be made up of 180 samples. All of samples are scanned at 96 dpi resolution in gray scale format which converted to the binary format with a constant threshold value. Some sample images are shown in Fig.3.
The extraction of scale invariant features is an essential problem in handwritten recognition. Major feature extraction methods that used for handwritten recognition have a high dimension that cause to increase the computational load. This paper proposed a scale invariant method based on gradient for feature extraction with low feature dimension. The proposed algorithm was tested by the 1-nearest neighbor (1-NN) and multi-layer perceptron (MLP). MLP is trained for each feature vector by using the parameters that redounded to best results.

As shown in Fig.4 the proposed method have better performance than the other approaches for two classifiers. MLP neural network is one of the most common families of neural networks for pattern classification tasks that belong to the class of supervised neural classifiers [42]. In this paper the MLP is used as the base classifier with one hidden layer that it is trained using the back–propagation algorithm minimizing the squared error criterion. The MLPs of all experts had 32 input nodes because of dimension of feature vector and 30 output nodes corresponding to 30 classes. The MLP has learning parameters, such as number of epochs, learning rate and number of neurons in hidden layer. The required number of epochs for reaching the highest recognition rate estimated by fourfold cross validation on the training set.

To find the sufficient number of neurons in hidden layer of experts, we accomplished different experiments with different number of neurons for 20 times. Fig.5 represents recognition rate of the neural network versus increasing the number of neurons in hidden layer. As shown in Fig.5 with more than 40 neurons in hidden layer experts reveals the best performance.
For diversifying base classifiers, the weights of MLP neural networks are initially set to small random values. In addition, different topologies are assumed for base classifiers with selecting different numbers of neurons for hidden layer and various learning rates. Table I shows the best topologies for base classifiers to be diverse and have highest recognition rate. Notice that experiment has been repeated 10 times to getting the recognition rates.

Table 1. The structure of base classifiers to have best results.

<table>
<thead>
<tr>
<th>Neurons in hidden layer</th>
<th>Eta</th>
<th>Epochs</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert1</td>
<td>45</td>
<td>0.1</td>
<td>400</td>
</tr>
<tr>
<td>Expert2</td>
<td>45</td>
<td>0.15</td>
<td>400</td>
</tr>
<tr>
<td>Expert3</td>
<td>47</td>
<td>0.2</td>
<td>400</td>
</tr>
<tr>
<td>Expert4</td>
<td>50</td>
<td>0.1</td>
<td>400</td>
</tr>
<tr>
<td>Expert5</td>
<td>50</td>
<td>0.11</td>
<td>400</td>
</tr>
</tbody>
</table>

To evaluate the performance of DTs as fusion classifier method in Farsi handwritten word recognition problem, it is compared with other fusion methods such as Maximum, Minimum, Median, Product and Average aggregation rules on gathered dataset. Experiments implemented with 3, 4 and 5 base classifier and results are represented in Table II. It can be seen from this Table that the best recognition rate of system for our database is 91.55. In DTs method the Euclidean distance is used to make decision about similarity between each of test sample and its corresponding class.

Table 2. Classification results(in %) of some fusion methods with 3, 4 and 5 base classifier.

<table>
<thead>
<tr>
<th></th>
<th>Max</th>
<th>Min</th>
<th>Median</th>
<th>Average</th>
<th>Product</th>
<th>DTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3expert</td>
<td>83.38</td>
<td>81.88</td>
<td>84.49</td>
<td>85.55</td>
<td>85.5</td>
<td>86</td>
</tr>
<tr>
<td>4expert</td>
<td>85.94</td>
<td>81.38</td>
<td>88.27</td>
<td>87.82</td>
<td>87.5</td>
<td>88.22</td>
</tr>
<tr>
<td>5expert</td>
<td>85.49</td>
<td>82.32</td>
<td>87.94</td>
<td>88.44</td>
<td>88.16</td>
<td>91.55</td>
</tr>
</tbody>
</table>
The experimental results show with fusion methods the performance of individual networks is significantly improved. To present how the errors are distributed over classes, which shows the level of classifiers diversity, we form a confusion matrix using the testing data. A confusion matrix [43] used to summarize results of a supervised classification. Performance of such systems is commonly evaluated using the data in the matrix. Fig.6 illustrates a graphic representation of the confusion matrices of each expert. Note that on each figure, symptoms on the main diagonal show correct classifications and others show misclassified samples.

![Figure 6. a-e) Confusion matrix of each expert in fusion. f) Confusion matrix of combined model. It is shown that recognition rate for combining of output of experts is considerably improved.](image)

5. Conclusions

This paper presented a new feature extraction method for Farsi handwritten word recognition. The proposed method is based on gradient, solves the problem of scale invariance, and it has a low feature dimension. This new feature extraction method and
related features, such as Zoning, Histogram, Gradient, Cross-Count and Wavelet transform, were tested by the 1-nearest neighbor (1-NN) and multi-layer perceptron (MLP). Decision Templates (DTs) introduced for classifier fusion. In order to evaluate the performance of a number of commonly used fusion techniques and Decision Templates, experiments with three, four and five base classifier are performed. Results exhibit that the DTs fusion model shows superior performance to the other techniques. We conduct experiments on Iranshahr dataset and validate the proposed algorithm with 91.55% of classification accuracy.

Acknowledgements
The authors wish to thank Iran Telecommunication Research Center (ITRC) for funding this project under number 500/8974.

References

Authors

Reza Ebrahimpour is an Assistant Professor at the Department of Electrical and Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran. He is also a research scientist at School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran. He received the B.S. degree in Electronics Engineering from Mazandaran University, Mazandaran, Iran and the M.S. degree in Biomedical Engineering from Tarbiat Modares University, Tehran, Iran, in 1999 and 2001, respectively. He obtained Ph.D. degree in the field of Computational Neuroscience from the School of Cognitive Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran in July 2007. Dr. Ebrahimpour is the author or co-author of more than 60 international journal and conference publications in his research areas, which include Human and Machine Vision, Computational Neuroscience, Neural Networks, Pattern Recognition and Multiple Classifier Systems.

Roya Davoudi Vahid was born in Hamedan, Iran, in Jan 1986. She received the BS degree in electronic engineering from Shahid Rajaee University, Tehran, Iran, in 2008. She is student in MS degree of Electronic Engineering in Shahid Beheshti University, Tehran, Iran. Her research interests include human and machine vision, neural networks and pattern recognition.

Babak Mazloom-Nezhad Maybodi was born in Tehran, Iran, in September 1960. He received the BS and Ms Degree in electronics engineering from Amir Kabir University, Tehran, Iran in 1985 and 1989, respectively. He received his PhD Degree in February 2008 from the Shahid Beheshti University, Institute of electrical and computer engineering, where he worked on low power CMOS mixed mode full adders. His research interests include digital signature, cryptography and E-commerce.