International Journal of Advanced Research in Engineering and Applied Sciences
ISSN: 2278-6252

PRESENTING A METHOD BASED ON STRUCTURAL FEATURES AND DECISION TREE FOR RECOGNITION OF PERSIAN HANDWRITTEN DIGITS

Elhan Nejad-Ababaf*

Navid Samimi*

Abstract: *What we have done is based extraction a series of structural features from a collection of handwritten digits. These features include: closed space on the number, branching and end points for half circle and the pixel density in different areas. In following decision tree created over these features and have been assessed.*

Why use a decision tree depends on optimal partitioning the samples in different categories and low time complexity.

Keywords: *Handwritten recognition, HODA dataset, fingerprint, decision tree, feature extraction, structural features*

*Department of Computer Engineering, Behbahan Branch, Islamic Azad University, Behbahan, Iran

1- INTRODUCTION

Some Part of the proposed method, is a sense of individual identity based on fingerprint. In the identification helping fingerprints, to minimize data about a fingerprint in the database, all files are not maintained completely. First, the entire image is analyze And then the key is saved. This plays an important role for fast searching in database. In general, about 35 important features exist in fingerprint image Such as crossing points, end points, branch etc. In Figure 1, a number of these points are illustrated#1Basically fingerprint recognition techniques are based on the relative position of the bifurcation points, bumps and Delta of each fingerprint and how far apart they may act. These characteristics are different in various types of fingerprints [1].

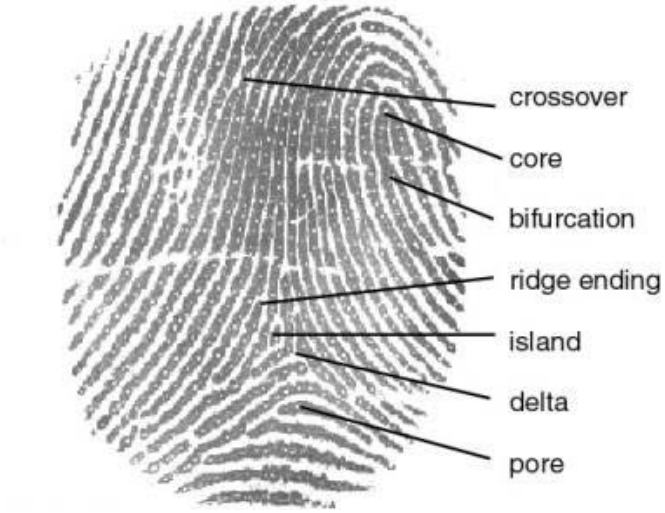

Figure 1: Characteristics are used for fingerprint recognition

In following we bring several clear characteristic of the decision tree:

- ✓ Decision tree gives us the ability to provide our predicts in the form of a set of rules.
- ✓ Decision tree do not requires very complex calculations to handle data classified.
- ✓ Decision trees are practical for different types of data such as continuous and category class of data.
- ✓ The decision tree shows us, which field or variable have dramatic impact on our prediction and packaging.

2- FEATURES OF THE PROPOSED METHOD

2-1- not sensitive to noise and curvature in drawing figures

Generally, Common weakness of many methods of feature extraction, are Lack of efficiency in facing with noisy figures. Different types of noises are: Spotter which likes dot, discrete line segments, Deformation, corrosion figures and some of the curvature [2]. Of course, some of these noises are before correcting in, pre- processing or post-processing.

2-2- Rely on structural and simple features

In this study Recognition of Persian handwritten digits has been done by using structural feature. These features have been selected Intuitive and empirical in a way that can be trusted element of geometrical structure and are independent from the author's writing style. Furthermore, understanding the proposed method for the audience to improve future work is very easy. While, in this way finding the new features and add them to plan can be accomplished to improve the recognition rate with minimal difficulty.

2-3- Simplicity of implementation

As we know, Understanding and implementation of the decision tree is very simple, due to it is consists of a series of if-then rules and obvious logic rules also. Rules Optimization for appropriate categories of cases, usually done in two ways; which are: optimization by help of programming and optimization by help of expert assistance.

2-4- Using by hybrid method

In many of the presented studies, Decision trees are used, as part of a combined approach to increase the detection rate [3]. For example, in handwritten character recognition in some works, firstly we use decision trees for separating dotted characters from un dotted ones. And Nervation is applied in order to continue detection. So, at least, the thing that can be done by help of the decision tree is to reduced n software of samples in to m (m << n). in presented research The decision tree is used alone, Because there are only 10 classes. However, in this study, the possibility of using the decision tree combined with other methods, is provided. It is essential to note, using another way along decision trees Will probably results better results and higher time complexity.

3- PRE-PROCESSING

As already mentioned, the proposed method do not required many of pre-processed (normalized) like normalizing Cross-condensation, normalizing curvature characters, and

normalizing the size. In other words, Extracted features are “reliable”. Pre-processing stage involves extracting figure skeleton (thinning). The only used preprocessing, is thinning the manuscripts numbers. In this section, the following code fragment used for thinning and obtain the skeleton of each digit [4].

4- FEATURE EXTRACTION

One of the most effective things, which can be resulted on Remarkable improvement of accuracy in a handwritten digits recognizing system is using appropriate features for the data representation. It requires choosing a suitable method for feature extraction and optimal Determination its parameters [5]. This section examines How to extract features, and important parameter in handwritten digits. The imported thing is extracting the features which can done most categorizing of samples. However, more features will lead to increase time complexity. Used Features can be divided into four categories:

- ✓ Closed space in the number
- ✓ End points and branching
- ✓ Semicircular direction
- ✓ The place of despite pixel density

5- CLASSIFICATION PHASE

At different stages, the classification features in Trees are chosen in such a way that we can provide the most possible classification, regarding the diversity manuscript and figures noise [6]. To illustrate the method, in Figure 2 you can see the Pseudo code of the proposed method.

```
numberbpixel = number of white pixels that have been surrounded by black pixels;
if(number_pixel > 20)
{
	remove all rows from bottom of digit that have only black pixel;
	number_removal = number of black pixel that have been deleted;
	if(number_removal < 8)
		if(number_pixel < 30)
			digit --> '0';
		else
			digit --> '5';
	else
	{
		discharge the pixels from eliminate state;
		if(direction of semi_circular == 'left')
			digit --> '6';
		else
			digit --> '9';
	}
}
else
{
	if(there are the only two end_points in up_side or down_side == true)
		if(end_points are being in the upper half == true)
			digit --> '7';
		else
			digit --> '8';
	else
		if(it has two semi_circular that one of them are in right_side and another is in the left_side == true)
			digit --> '4';
		else
		{
			remove all rows from buttom of digit that have only black pixl;
			if(there is a a semi_circular to the right side == true)
				digit --> '4';
			else
			{
				if(is there any semi_circular toward up == true)
					if(are there semi_circular toward up and down == true)
						digit --> '3';
				else
				{
					If(are there two branch_point and two semi_circular toward up == true)
						digit --> '3';
					else
						if(is there only one semi_circular toward left == true)
							digit --> '6';
						else
							digit --> '2';

				}
					discharge the pixels from eliminate state;
					if(concentration in upper half >> lower half)
						if(black pixel it right half > 2 * black pixel in left half)
							digit --> '6';
						else
							digit --> '2';
					else
						 if(black pixel > 20)
							digit --> '1';
						else
							digit --> '0';
			}

		}
}
```

Figure 2: Pseudo code of the proposed method

6- RESULT OF PROGRAM RUNNING

The program was implemented using the MATLAB software. At any stage, the MATLAB Tool Box does not used. And coding was implemented at all stages. At diagnosis phase, 100 samples were selected randomly from each class. The results can be seen in Table 1.

Table 1: handwritten digits recognition rate of the proposed system

digit	The number of samples correctly diagnosed	The number of samples wrongly diagnosed	Percentage of correct diagnosis
0	88	12	88%
1	96	4	96%
2	92	8	92%
3	87	13	87%
4	90	10	90%
5	89	11	89%
6	88	12	88%
7	100	0	100%
8	99	1	99%
9	92	8	92%
total	921	79	92.1%

Similarly, the code running results in false and true detection. And that, if improperly diagnosed, Attributed to each class, can be seen in Table 2.

Table 2: Results of the implementation of the Code

	0	1	2	3	4	5	6	7	8	9
0	88	3	0	0	0	9	0	0	0	0
1	2	96	0	0	1	0	1	0	0	0
2	0	0	92	3	1	0	4	0	0	0
3	0	0	8	87	3	0	2	0	0	0
4	0	0	3	2	90	1	4	0	0	0
5	8	0	0	0	0	89	1	0	0	2
6	0	2	6	1	0	0	88	0	0	3
7	0	0	0	0	0	0	0	100	0	0
8	0	0	0	0	0	1	0	0	99	0
9	1	0	0	0	0	2	5	0	0	92

7- CONCLUSION

In this study, a new method based on the new features to recognize handwritten Persian digits, was offered. Although there are some challenges to some figures, detection rate is acceptable. The proposed method can be used in applications, with some modifications. Easily implemented and lower time complexity is One of Strength point of the proposed method in detection Phase.

RFERENCES

[1] Sadri, J., Suen, C.Y., Bui, T.D., "A genetic Framework using contextual knowledge for segmentation and recognition of handwritten numeral strings", Journal of Pattern Recognition, Elsevier, 2006.

[2] Arica N. and Yarman-Vural F.T., "One dimensional representation of two dimensional information for HMM based handwritten recognition", Pattern Recognit. Lett., vol. 21, No. 6–7, pp. 583–592, 2009

[3] Arica Nazif ,Yamin-Vural Fatos T., "An overview of character recognition based focused on off-line handwriting", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 31, No. 2, May 2007.

[4] Hu Peifeng, Zhao Yannan, Yang Zehong, Wang Jiaqin, "Recognition of gray character using gabor filters", Proceedings of the Fifth International Conference on Information Fusion, Vol.1 , pp. 419 -424, July 2002.

[5] Soltanzadeh, H., Rahmati, M., "Recognition of Persian handwritten digits using image profiles of multiple orientations", Journal of Pattern Recognition, Elsevier, 2004.

[6] Azmi R., Kabir E., "A recognition algorithm for hand-printed farsi characters", Proc. Of ICT96, Istanbul, Turkey, pp. 852-855, 1996.