

Handwritten Digit Recognition Using Multiple Feature Extraction Techniques and Classifier Ensemble

Rafael M. O. Cruz, George D. C. Cavalcanti and Tsang Ing Ren
 Center of Informatics, Federal University of Pernambuco
 www.cin.ufpe.br/~viisar
 Recife, Brazil
 {rmoc,gdcc,tir}@cin.ufpe.br

Abstract— It is herein proposed a handwritten digit recognition system which uses multiple feature extraction methods and classifier ensemble. The combination of the feature extraction methods is motivated by the observation that different feature extraction algorithms have a better discriminative power for some types of digits. Six features sets were extracted, two proposed by the authors and four published in previous works. It is shown that combining these feature sets is sufficient to achieve high recognition rates. Several combination schemes were tested, showing good results. A scheme using neural networks as a combiner achieved a recognition rate of 99.68%, the highest one on the MNIST database.

Keywords—*Handwritten recognition, Feature Extraction, Classifier Ensemble, Neural Networks.*

I. INTRODUCTION

Unconstrained handwritten digit recognition is one of the most important problems in computer vision. There is a great interest in this area due to many potential applications, especially where a large number of documents must be analyzed, such as post mail sorting, bank check analysis and handwritten forms processing. Many approaches have been proposed with high recognition rates recently [1, 2, 3, 4, 5, 6], however there is still room to increase the recognition accuracy because an error can be very costly in some applications.

In this paper, a novel recognition system by using combination of multiple feature extraction methods and an ensemble classifier system is proposed. Six feature set using different approaches (projections, zoning, edges, concavities and gradient) were extracted, Two of them are proposed in this work while four were extracted from previous works, to add diversity to the system. The diversity in the methods is important as some methods have a better ability to recognize certain types of images. This paper shows that using the six feature extraction methods, none pattern presented error on all of them. Therefore these six methods are sufficient to achieve a very high recognition rate. The problem now is to find the best scheme to combine them.

The six feature sets are trained by Multi-Layer Perceptron (MLP) neural networks separately. Their outputs are then combined to give a more accurate output. Several combination schemes have been tested on the well-known MNIST handwritten digits database. The results show that all combination schemes greatly improved the recognition performance when compared to a single feature extraction-classifier pair alone. A combination module using another MLP network as combiner is proposed, achieving a recognition rate of 99.68% on the MNIST database, which is the highest recognition rate published for this database to date.

This paper is organized as follow: In Section II the six feature extraction algorithms are briefly introduced. In Section III the results obtained by each feature set is analyzed. The ensemble classifier system and its results are presented in Section IV. The conclusion is given in the final section.

II. FEATURE EXTRACTION

A total of six feature extraction algorithms were used. The methods Multi Zoning and the Modified Edge Maps are proposed in this paper while four other methods, Structural Characteristics, Projections, Concavities Measurements and Gradient Directional were proposed in previous works. The methods are described below.

A. Structural Characteristics

This algorithm consists in extracting histograms and profiles, combining then in a single feature vector. The input image is scaled in a 32 x 32 matrix. Horizontal and vertical histograms are computed by the number of black pixels in each line and column, respectively. Radial histogram is computed by the number of black pixels in 72 directions at 5 degree intervals. Radial In-Out and Radial Out-In profiles are calculated by the position of the first and the last black pixel, respectively, that starts from the center and goes to the border in 72 directions at 5 degree intervals. These features form a 280-dimension $(32 + 32 + 72 + 72 + 72)$ feature vector. Details can be found in [7].

B. Modified Edge Maps

An $N \times N$ image is thinned and scaled into a 25×25 matrix. The Sobel operators are used to extract four distinct edge maps: horizontal, vertical and two diagonals (45° and -45°). These four maps and the original image are divided into 25 sub-images of 5×5 pixels each. The features are obtained calculating the percentage of black pixels in each sub-image (25 features per image). The features are combined to form a single feature vector containing 125 (25×5) features.

C. Image Projections

This method consists of extracting the radial and diagonal projections. To extract the radial projections, the image must first be divided into four quadrants: top, bottom, right and left. The quadrants are used to remove rotational invariance which is clearly undesirable in handwritten digit recognition. Radial projections are obtained by grouping pixels by its radial distance to the center of the image in each quadrant separately. The diagonal projection is computed simply by grouping pixels by the two diagonal lines (45° and -45°). More details can be found in [8]. The values of each projection are normalized to a range $[0; 1]$ through the division by the maximum value. The normalized features are concatenated in a single vector containing 128 features (16 for each radial projection and 32 for each diagonal projection).

D. Multi Zoning

In this algorithm, an $N \times N$ character image is divided into several sub-images and the percentage of black pixels in each sub-image is used as feature. To achieve better recognition performance, many different configurations of division were selected and concatenated to form the feature vector. A total of 13 different configurations (3×1 , 1×3 , 2×3 , 3×2 , 3×3 , 1×4 , 4×1 , 4×4 , 6×1 , 1×6 , 6×2 , 2×6 and 6×6) were chosen, resulting in $123(3 + 3 + 6 + 6 + 9 + 4 + 4 + 16 + 6 + 6 + 18 + 18 + 36)$ features.

E. Concavities Measurement

The following steps are used to measure the concavities. Firstly the image is scaled into a 18×15 matrix. The image is divided in six zones, each one containing its own 13-dimension feature vector. Each position of the feature vector corresponds to one of the possible configurations (i.e. number of black pixels reached and its positions). For each white pixel, the algorithm search in four directions the number of black pixels that it can reach and the directions that a black pixel is not reached. The position on the feature vector which is related to the configuration found in the search is incremented. The feature vectors of each zone are combined in a single vector with 78 (13×6) features. A detailed version of the algorithm can be found in [9].

F. MAT-Based Gradient Directional Features

This algorithm computes the gradient components in a grayscale image. Grayscale images are used because they have richer information than a binary image for discrimination [2]. Thus, before start the algorithm, a binary input image is first transformed in a pseudo-

grayscale one using the Medial Axial Transformation (MAT) algorithm. The Sobel operators are used to generate the gradient amplitude and phases. The gradient magnitude of each pixel is quantized into one of eight directions at $\pi/4$ intervals. The image is divided into 16 sub-images and for each sub-image, the number of pixels in each of the eight directions is computed as feature. The feature vector size is 128 (16 sub-images \times 8 directions). Details can be found in [2].

III. EXPERIMENTS AND RESULTS

This section presents the experiments obtained by each feature extraction method separately. All the experiments were conducted on the well-known MNIST¹ database. This database contains a training set of 60,000 images and a test set of 10,000 images. All digits are size-normalized and centered in a 28×28 image. The training set was divided in 50,000 patterns for training (5,000 images per digit) and 10,000 (1,000 per digit) for validation. A three layers MLP trained with Resilient Backpropagation [10] algorithm was used as classifier in all feature sets.

After preliminary tests, the best configurations for each feature set were selected. For the methods Edge Maps and Gradient Directional the number of nodes in the hidden layer was 300. For the methods Zoning, Structural Characteristics, Concavities Measurement and Image Projections the number of nodes in the hidden layer was 360, 340, 175 and 330 respectively. The best results for each feature set are shown in Table I.

TABLE I. BEST RESULTS FOR EACH FEATURE EXTRACTION METHOD

Digit	structural	edge	projections	Zoning	Concavities	Gradient
0	1.12%	2.14%	1.83%	1.12%	3.87%	2.04%
1	0.88%	1.85%	1.58%	1.05%	1.67%	1.32%
2	3.97%	4.74%	4.74%	3.77%	4.34%	4.84%
3	3.86%	5.24%	5.24%	3.16%	8.31%	5.54%
4	2.75%	7.85%	3.67%	2.95%	7.02%	3.06%
5	4.37%	5.27%	6.39%	3.03%	4.44%	3.70%
6	2.19%	3.34%	2.82%	2.92%	3.65%	2.61%
7	3.11%	6.23%	4.57%	4.38%	5.62%	4.96%
8	4.00%	6.46%	6.26%	4.10%	10.36%	6.46%
9	4.40%	9.42%	6.15%	4.86%	7.98%	7.34%
Mean	3.05%	5.22%	4.28%	3.12%	5.69%	4.17%

It can be seen that different feature extraction methods have a better discriminative power for certain classes of digits. It is important to observe that even for the same digits, the errors made in different feature sets are different. In Figure 1 the intersection of errors using 3 feature sets is shown. Only 9 patterns were misclassified on these 3 feature sets. Using the 6 feature extraction algorithms, 1560 different digits presented errors (error made by more than one method is counted as one error). The majority of errors are made by only one feature extraction method (1026) while none are made by all of them. The numbers of samples that are misclassified per number of methods are shown in Table 2. The number of errors made by only one method (any

¹ <http://yann.lecun.com/exdb/mnist/>

method) is shown in the first column. The numbers of errors made by only two methods (any two) is shown in the second column and so forth. These numbers show the information of these six features sets are really complementary and also prove the fact that some feature extraction algorithms perform better for some images. This can be explained by the diversity on the feature extraction algorithms.

It is also important to observe that none pattern were misclassified by all feature sets, therefore an ideal combination of these six techniques would achieve a 100% recognition rate for the MNIST database. For this reason, these six methods were chosen to create the ensemble system. The problem now is to find the best combination scheme for this given task.

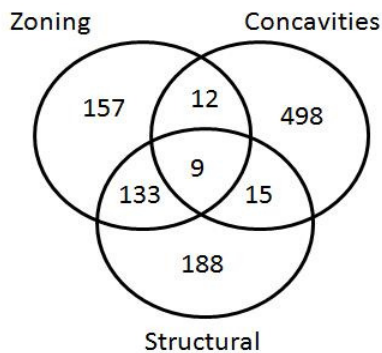


Figure 1. Intersection of errors obtained by the feature sets Structural Characteristics, Zoning and Concavities Measurement

TABLE II. DISTRIBUTION OF ERRORS BETWEEN METHODS

No. of methods which made the same error	1	2	3	4	5	6	Total
No. errors	1026	223	119	188	4	0	1560

IV. ENSEMBLE SYSTEM

Combination of classifiers has been widely studied in past years as an alternative to increase efficiency and accuracy [11]. The main motivation for using classifier ensemble in the given task comes from the observation that the errors made by the classifiers with different feature extraction methods does not overlap. Another motivation comes from the divide and conquer paradigm (i.e., using each feature extraction method separately and combining their results instead of use a single set consisting of the six feature extraction methods).

A diagram of the ensemble system is shown in Figure 2. The six feature extraction methods were used on the system because of the observation they alone can lead to a perfect recognition rate. The ensemble system consists of the six feature extraction techniques using the same MLP configuration used in the experiments above and a combination module. each MLP network estimate the posteriori probability for each digit and send them to the combination module.

Both fixed and trained combination rules were used. The fixed combination rules used were: Sum, Product,

Maximum, Median and Voting. The theoretical framework for fixed combination rules can be found in [11]. For the trained combination rule, a MLP network with one hidden layer was used. Training combiner usually presents better recognition rates as the combiner can adapt itself to the classification problem [12]. The MLP combiner was trained selecting 50.000 images of the training set (5000 per digit) for training and 10.000 images (1000 per digit) for validation. For each image, the posteriori probability are estimated by each feature extraction method and used as features to the network. The Resilient Backpropagation algorithm [10] was used to train the network and the number of nodes in the hidden layer was set as 50. The results of each combination rule are shown in Table 3.

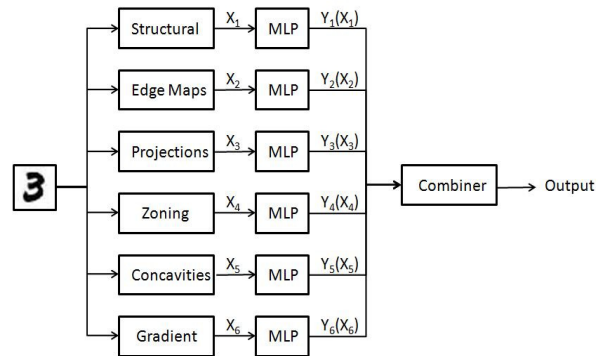


Figure 2. Classifier Ensemble System

It can be seen that all combination rules showed great improvements in relation to all feature extraction-classifier pair alone. The module using a MLP network as combiner obtained an error rate of 0.32% which was the best result. This can be explained by the ability of the network in learning how to perform the best combination using the training set. The Maximum rule also presented a good result. This can be explained by the ability some of these feature extraction methods have to recognize certain types of digits.

TABLE III. BEST RESULTS FOR EACH COMBINATION RULE

Digit	sum	product	Max	Median	Voting	MLP
0	0.20%	0.61%	0.00%	0.30%	0.40%	0.10%
1	0.35%	0.44%	0.00%	0.35%	0.35%	0.08%
2	0.96%	1.25%	0.09%	1.60%	1.16%	0.19%
3	1.38%	0.99%	0.69%	1.08%	1.38%	0.39%
4	1.02%	0.71%	0.30%	1.22%	1.12%	0.30%
5	1.12%	0.89%	0.89%	1.57%	2.24%	0.67%
6	1.46%	1.14%	0.20%	1.35%	1.77%	0.20%
7	1.26%	1.46%	1.07%	2.04%	2.23%	0.48%
8	1.33%	0.92%	1.33%	1.54%	2.15%	0.20%
9	2.57%	2.18%	1.88%	2.28%	4.06%	0.59%
Mean	1.16%	1.06%	0.64%	1.27%	1.67%	0.32%

The best results obtained in the last years for the MNIST database are shown in Table 4. The proposed combination scheme outperformed all the previous results on this database. It is also important to observe

that many of the best results [1, 3, 4, 6] are based in Convolutional Neural Networks. They also need to increase the training set using distortions to create new patterns from the training images to achieve good recognition rates. Thus, this paper shows a different approach to achieve high performance in handwritten recognition and without the need of increase the training set by distortions.

The misclassified images by the proposed system are shown in the Figure 3. Based on these images, it can be seen that many of the misclassifications were made in digits that are ambiguous either by noise, distortion, segmentation problem or peculiar writing style. To increase the reliability of the system, a strategy to reject these ambiguous digits must be investigated. On the other hand, some of the misclassified digits can be easily recognized by humans. Therefore, the recognition rate in this database can still be improved. Some improvements might be done by adding another feature extraction method to the system, especially a method with better discriminative power for the digits 9 and 5 as all the feature extraction techniques in this paper presented high error rates for these digits.

TABLE IV. COMPARATIVE RESULTS ON THE MNIST DATABASE

Method	Distortions	Error(%)
Boosted LeNet-4 [1]	Affine	0.70
TfE-SVM [3]	Affine	0.44
PNCN Classifier [5]	Skewing	0.44
Cascade Ensemble Classifier (Without Rejection) [2]	-	0.41
Convolutional Neural Net. [4]	Elastic	0.40
Large Conv. Net + Unsup. pre-training [6]	Elastic	0.39
Proposed	-	0.32

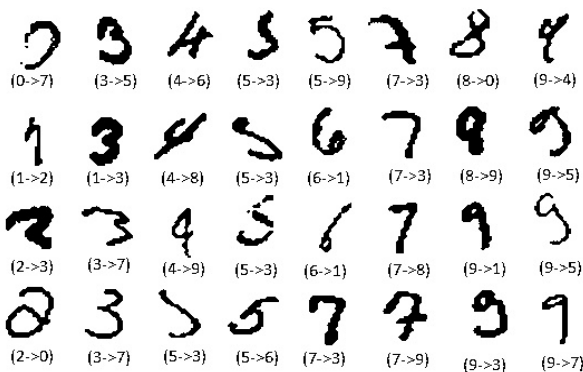


Figure 3. List of Misclassified images using the MLP combination scheme

V. CONCLUSION

In this paper, a method to increase handwritten digits recognition rates by combining feature extractions methods is proposed. Six feature sets, two proposed in

this paper, using different approaches were extracted and evaluated. The experiments demonstrated that different feature extraction algorithms, have better discriminative ability for certain types digits, therefore they are complementary.

Based on these experiments a classifier ensemble consisting of the six feature sets presented was proposed. Some of the most common combination rules were evaluated as well as one trained rule. The results showed all combination rule greatly improved the recognition performance. The experiment using a MLP as a trained combiner achieve a recognition a error rate of 0.32% outperforming previous results published on the MNIST database.

Some of the misclassified digits are ambiguous either by segmentation problems, peculiar writing style or distortions. A strategy to reject this ambiguous digits and improvements to recognize the misclassified digits that can be easily recognized by humans are being studied.

REFERENCES

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] Ping Zhang, *Reliable recognition of handwritten digits using a cascade ensemble classifier system and hybrid features*, Ph.D. thesis, Concordia University, Montreal, P.Q., Canada, 2006.
- [3] Fabien Lauer, Ching Y. Suen, and Gérard Bloch, "A trainable feature extractor for handwritten digit recognition," *Pattern Recognition*, vol. 40, no. 6, pp. 1816–1824, 2007.
- [4] Patrice Y. Simard, Dave Steinkraus, and John C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," *International Conference on Document Analysis and Recognition*, vol. 2, pp. 958–963, 2003.
- [5] E.M. Kussul, T.N. Baidyk, D.C. Wunsch, O. Makeyev, and A. Martin, "Permutation coding technique for image recognition systems," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1566–1579, 2006.
- [6] Marc'Aurelio Ranzato, Y-Lan Boureau, and Yann Le-Cun, "Sparse feature learning for deep belief networks," in *Advances in Neural Information Processing Systems*, 2007.
- [7] E. Kavallieratou, K. Sgarbas, N. Fakotakis, and G. Kokkinakis, "Handwritten word recognition based on structural characteristics and lexical support," *International Conference on Document Analysis and Recognition*, pp. 562–567, 2003.
- [8] Y.C. Chim, A. A. Kassim, and Y. Ibrahim, "Dual classifier system for handprinted alphanumeric character recognition," *Pattern Analysis and Application*, no. 1, pp. 155–162, 1998.
- [9] Luiz S. Oliveira, Robert Sabourin, Flávio Bortolozzi, and Ching Y. Suen, "Automatic recognition of handwritten numerical strings: A recognition and verification strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1438–1454, 2002.
- [10] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *International Conference on Neural Networks*, pp. 586–591, 1993.
- [11] Josef Kittler, Mohamad Hatef, Robert P. W. Duin, and Jiri Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226–239, 1998.
- [12] R. P. W. Duin, "The combining classifier: to train or not to train?," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, vol. 2, pp. 765–770 vol.2.