Decision Fusion and Reliability Control in Handwritten Digit Recognition System

Dusan Cakmakov, Vladimir Radevski*, Younes Bennani** and Dejan Gorgevik***

University "Sv. Kiril i Metodij", Department of Mathematics and Computer Science, Skopje, Macedonia
*University Galatasaray, Department of Computer Engineering, Ortakoy, Istanbul, Turkey
**University Paris 13, LIPN-CNRS, Villetaneuse, France
***University "Sv. Kiril i Metodij", Department of Computer and Information Technology, Skopje, Macedonia

In this paper, the cooperation of two feature families for handwritten digit recognition using a committee of Neural Network (NN) classifiers will be examined. Various cooperation schemes will be investigated and corresponding results will be presented. To improve the system reliability, we will upgrade the committee scheme using multistage classification based on rule-based and statistical cooperation. The rule-based cooperation enables an easy and efficient implementation of various rejection criteria while the statistical cooperation offers better possibility for fine-tuning of the recognition versus the reliability tradeoff. The final system has been implemented using rule-based reasoning with rejection criteria for classifier decision fusion and the generalized committee cooperation scheme for classification of the rejected digit patterns. The presented results show that we propose a successful approach for reliability control in committee classifier environment and indicate that a suitable cooperation of statistical and rule-based decision fusion is a promising approach in handwritten recognition systems.

Keywords: multistage classification, rejection, structural, statistical, features.

1. Introduction

The idea of combining classifiers in order to compensate their individual weakness and to preserve their individual strength has been widely used in recent pattern recognition applications (Xu et al., 1992), (Ho et al., 1994), (Kittler et al., 1998), (Duin and Tax, 2000), (Roli et al., 2001). Data from more than one source that are separately processed, can often be profitably re-combined to produce more concise, more complete and/or more accurate situation description. A theoretical and mathematical framework that explains the reasons for expecting the improvement of the performances in cases of combining classifier outputs can be found in (Kittler, 1998), (Tumer and Ghosh, 1999).

The classical paradigm for character recognition is concentrated around two steps, feature extraction, where an appropriate representation of the pattern is developed, and classification, where decision rules for separating pattern classes are defined. Combining features of different nature and the corresponding classifiers has been shown to be a promising approach in handwritten recognition systems (Kimura and Shridar, 1991), (Huang and Suen, 1993), (Suzuki et al.,1995), (Yamaguchi et al., 1997), (Dahmen et al., 2001).

In this paper, we present a two-stage classification system for handwritten digit recognition using a committee of NN classifiers. We start with two NN classifiers which work on two different feature families for the same digit image. Our feature families are referenced as structural and statistical feature sets (Radevski and Bennani, 2000), and they differ (especially structural features) from the feature sets with the same reference used in other handwritten recognition systems (Duer et al., 1980), (Heute et al., 1996), (Lou et al., 1999).

In order to improve the system reliability, we introduced two-stage classification based on rule-based and statistical cooperation. The rule-based cooperation enables an easy and efficient implementation of rejection criteria. The statistical cooperation offers better possibility for fine-tuning of the recognition versus the reliability tradeoff. The final system has been implemented using rule-based reasoning with rejection criteria for the classifier cooperation and the generalized committee cooperation scheme for classification of the rejected digit patterns. Our goals in this paper are to examine usefulness of our feature extraction and selection technique, to show an approach for reliability control in committee classifier environment and to present a two stage classifier cooperation using rule-based and statistical decision fusion rather than to compete with the recognition rates of other handwritten digit recognition systems (LeCun et al., 1995).

2. The System Architecture

The recognition system is constructed around a modular architecture of feature extraction and digit classification units. The preprocessed image is input for the feature extraction module, which transfers the extracted features toward NN classifiers (Fig. 1). From the digit images with resolution of 128 $\times$ 128 pixels, we obtained 16 $\times$ 16 binary images on which the smoothing and centralizing preprocessing techniques have been applied. We have extracted 54 structural and 62 statistical features. The structural and statistical feature sets are forwarded to two separate NN classifiers where the feature selection procedure is performed. NN classifier outputs are combined using rule-based and statistical cooperation. On this level, rejection criteria are introduced and the corresponding system reliabilities are calculated.

2.1. The Structural Features

The structural feature set is a domain-dependent set. Its nature and the techniques implemented for detection and extraction are strongly dependent of the nature of the objects to be recognized.

The first step in the creating of the structural feature set is defining a reasonable set of elementary shape primitives for digit constructions. We have proposed 27 elementary primitives showed in Fig. 2. The digit image is searched for these primitives twice: firstly on the original digit image orientation, and secondly on the rotated digit image for 90°. So, the total number of

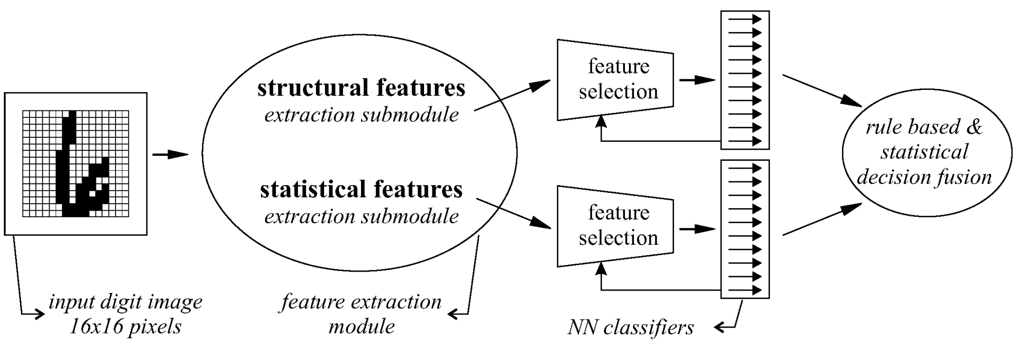


Fig. 1. The system architecture.

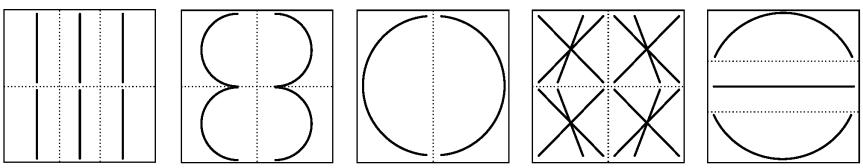


Fig. 2. Image sub-regions and the elementary primitives.

primitives is 54, and that is the number of the elements in the structural feature set.

The detection and extraction of the structural features is performed by dividing the image binary matrix into two, three, four and six sub-regions. The existing shape in each of those sub-regions is compared with the proposed primitives in the same sub-regions whose existence is expected.

We have made the search for the primitives with parallel lines from up, down, left or right side, depending on their position. In this way, we have obtained control points for the shape description (Fig. 3a). The shape is represented by the lengths of the line segments s_i (between the control points), and the corresponding angles between the line segments and the x-axis (Fig. 3b).

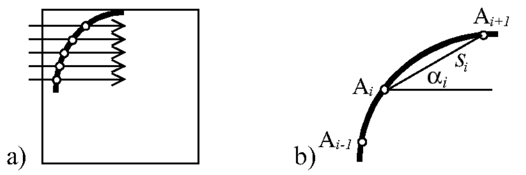


Fig. 3. a) Line control points;
b) Line representation: segments and angles.

The similarity measure between the found shape and the corresponding predefined primitive is based on the differences between the changes of angles along the shapes, normalized to take value between 0 and 1. This similarity measure is a simplified variation of the curve matching technique described in (Cakmakov, 1998).

Thus, the structural feature is composed of 54 values of the calculated similarities between the found shapes in the corresponding sub-regions and the corresponding elementary primitives.

2.2. The Statistical Features

The statistical feature set is composed of 62 features that give the pixel-based information presented by the densities of the lit pixels in various regions of the digit image. The first 54 statistical features are obtained from the projection histograms obtained by the vertical (16), horizontal (16) and two diagonal (22) projections (5 pixels left and right around the main diagonals). The last 8 features are obtained from the zone-pattern regions showed in Fig. 4.

Each of the numerical values of the 62 statistical features represents the filled up percentage of the projection histograms. So, the statistical features have values between 0 and 1. This kind of features in different forms has been exploited in many pattern recognition systems. Burel and colleagues called them oriented profiles (the projection histograms in our case), and statistical features (the zone-pattern features in our case) (Burel et al., 1992).

2.3. Feature Selection

In order to optimize member classifiers, we included a feature selection phase. Various feature selection techniques have been proposed (Liu and Motoda 1998). We have implemented

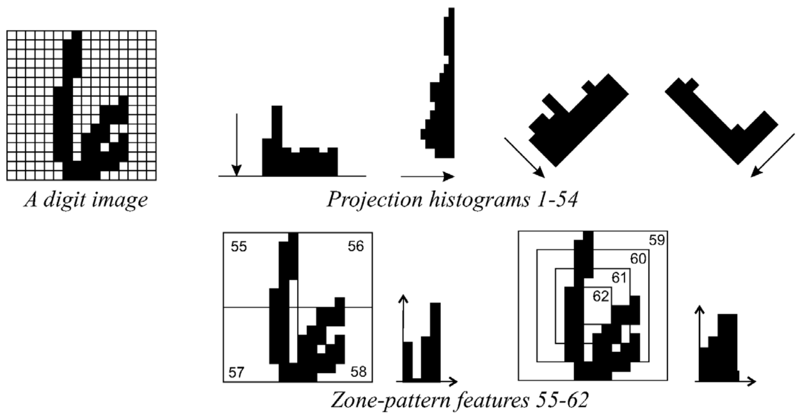


Fig. 4. Projection histograms and zone-pattern features.

a supervised feature selection based on the Optimal Cell Damage technique (Cibas et al. 1994). The features are selected one by one, according to the error rate of the classifier. The selection phase is deduced to deleting of the input NN nodes, after establishing a threshold value for their saliency. Thus, the feature selection phase prunes the input NN nodes with lower importance. This procedure is performed separately on the member classifiers. This reduces the system complexity by at least 15 to 20%, keeping the same recognition rate.

2.4. The Handwritten Digit Data Base

The database for our experiments is an extraction of the NIST (National Institute of Standards and Technology) handwritten digit database. The digit images are consisted of gray level pixels presented with real numbers in [-1, 1] interval. The total number of 23898 digit images is divided into two groups, 17952 images for the training phase and 5946 images for the test phase.

The digits from the original database are rearranged in order that digits in the test set belong to different writers from those in the learning set.

In Fig. 5, a fragment and the composition of the digit database are presented.

The class distribution of the samples in the learning and the test set is nearly uniform (Fig. 5).

3. Committee cooperation schemes

The cooperation schemes are designed around two full connected Multi-layer Perceptron (MLP) NNs with one hidden layer. They perform the classification task on the sets of structural and statistical features respectively. The size of their input layers corresponds to the cardinality of the corresponding feature set, and the number of hidden layer nodes is determined experimentally (Fig. 6). Each output node corresponds to one digit. The combined classification unit is based on different cooperation schemes.

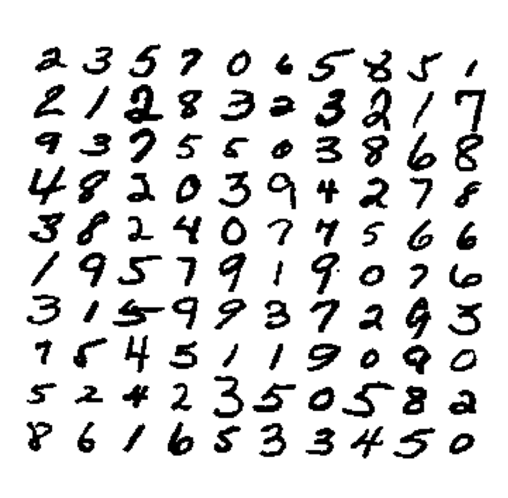
Let us $y_i(\boldsymbol{x})$, $i = 1, 2$ be the output for input feature vector $\boldsymbol{x}$ of two MLPs for structural and statistical features respectively. Using the individual classifier information, we have examined a few cooperation schemes (Barabas 1983), (Xu et al. 1992), (Ho et al. 1994), (Bishop 1995), (Kittler et al. 1998):

— Simple average

This simplest cooperation is based on the average of the rescaled outputs from the individual classifiers. In our case, we have

$$y_{ave}(\boldsymbol{x}) = \frac{1}{2}\sum_{k=1}^{2} y_k(\boldsymbol{x}).$$

This approach does not take into consideration the measure of the belief of each of the member decisions.



Class	Learning Set: 17952 samples	Test Set: 5946 samples
0	1860 (10.36%)	606 (10.19%)
1	2026 (11.29%)	670 (11.23%)
2	1750 (9.75%)	594 (9.99%)
3	1895 (10.56%)	622 (10.46%)
4	1714 (9.55%)	556 (9.35%)
5	1535 (8.55%)	515 (8.66%)
6	1726 (9.61%)	591 (9.94%)
7	1878 (10.46%)	613 (10.31%)
8	1783 (9.93%)	589 (9.91%)
9	1785 (9.94%)	590 (9.92%)

Fig. 5. A fragment and the composition of the digit database.

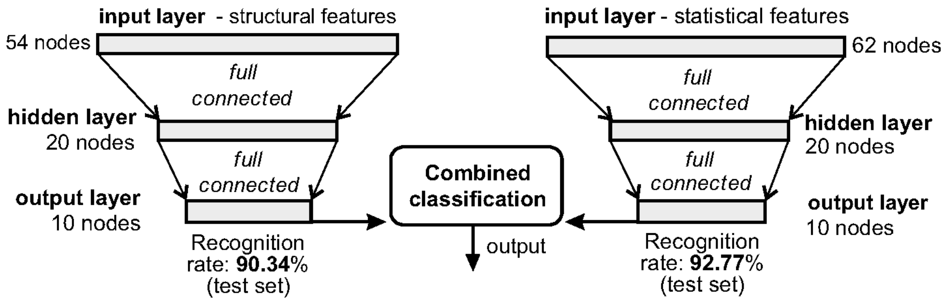


Fig. 6. The classification module.

— Classification task

In this cooperation, we consider the outputs of the individual classifiers as a data to be classified by another MLP NN, fully connected with one hidden layer. The input layer contains 20 units, 10 from the output of each individual classifier. The hidden layer contains 15 units and the output layers 10 units.

— Dempster's rule

Here, the probability of the favoring of the class C_i is computed as a product of the individual classifier outputs $y_k(\boldsymbol{x} \in \mathbf{C}_i)$

$$y_{demp}(\boldsymbol{x} \in \mathbf{C}_1) = \frac{1}{\eta}\Pi_{k=1}^{2} y_k(\boldsymbol{x} \in \mathbf{C}_i)$$

where normalization η is calculated by the formula

$$\eta = \sum_{i=1}^{10} \Pi_{k=1}^{2} y_k(\boldsymbol{x} \in \mathbf{C}_i).$$

This cooperation takes into account the fuzziness of the classifier votes, by giving less confidence to less certain votes.

— Bayesian formalism

Bayesian Formalism uses the error of each classifier, represented by the corresponding confusion matrices $[n_{i,j}^1]$ and $[n_{i,j}^2]$. The element means that the elements of the class C_i have been assigned to the class C_j by the classifier k. If we consider the confusion matrices as sources of prior knowledge, they can be used for estimation of the certainty for each classifier. With this knowledge, the conditional probability that the proposition $\boldsymbol{x} \in \mathbf{C}_i$, $i = 1, 10$ is true under the decision $D_k(\boldsymbol{x}) = C_j$, $k = 1, 2$ made by each of the classifiers is calculated by

$$P(\boldsymbol{x} \in \mathbf{C}_i / D_k(\boldsymbol{x}) = \mathbf{C}_j) = \frac{n_{i,j}^k}{\sum_{i=1}^{10} n_{i,j}^k}.$$

So, the final classification is made by the formula

$$P_{bayes}(\boldsymbol{x} \in \mathbf{C}_i) = \frac{1}{\eta}\Pi_{k=1}^{2} P(\boldsymbol{x} \in \mathbf{C}_i / D_k(\boldsymbol{x}) = \mathbf{C}_j),$$

where normalization η is calculated by

$$\eta = \sum_{i=1}^{10} \Pi_{k=1}^{2} P(\boldsymbol{x} \in \mathbf{C}_i / D_k(\boldsymbol{x}) = \mathbf{C}_j).$$

— Generalized committee cooperation scheme

Generalized committee prediction is given by a weighted combination of the predictions of the member classifiers

$$y_{gen}(\boldsymbol{x}) = \sum_{i=1}^{2} \gamma_i y_i(\boldsymbol{x}).$$

For the finite-sample approximation of the correlation matrix M, we can obtain the solutions for the parameters γ_i

$$M_{i,j} \approx \frac{1}{N}\sum_{n=1}^{N} (y_i(\boldsymbol{x}_n) - \mathbf{t}_n)(y_j(\boldsymbol{x}_n) - \mathbf{t}_n)$$

and

$$\gamma_i = \frac{\sum_{j=1}^{2} (M^{-1})_{i,j}}{\sum_{k=1}^{2}\sum_{j=1}^{2} (M^{-1})_{k,j}},$$

where $\mathbf{t}_n$ is the target value corresponding to the input vector $\boldsymbol{x}_n$, and N is the number of examples in the training set. Following this procedure we have obtained $\gamma_1 = 0.37$ and $\gamma_2 = 0.63$. So, the committee act is performed according to

$$y_{gen}(\boldsymbol{x}) = 0.37 y_1(\boldsymbol{x}) + 0.63 y_2(\boldsymbol{x}).$$

The recognition results of the cooperation of two classifiers according to the described cooperation schemes are given in Table 1.

Committee decision rule	Recognition rate Test set (%)
Simple average	94.97
Classification task	95.22
Dempster's rule	94.87
Bayesian formalism	95.04
Generalized committee	95.09

Table 1. The results of the classifier cooperation.

The cooperation results presented in Table 1 show that the non-linear combination of the classifier outputs given by the classification task and the generalized committee provide better recognition results. However, we find that the misclassification rate of approximately 5% is not satisfactory.

To design a recognition system with higher reliability, we will study the possibilities of introducing more sophisticated multistage cooperation including the implementation of rejection criteria (Radevski and Bennani 2000).

4. On the individual classification units

In order to establish more effective cooperation scheme, we have studied the properties of the individual classifiers.

4.1. Analysis of classifier outputs

Recognition rates of the NN classifiers and the recognition rates of the "top two" cases (the correct decision is the first or the second classifier choice) are evaluated and given in Table 2.

Classification module	Recognition rate Test set (%)	"top two" recognition rate Test set (%)
structural features	90.34	95.52
statistical features	92.77	97.00

Table 2. Recognition rates of the individual classifiers.

The high recognition rates in the "top two" cases show that the number of right answers among the first two choices of the classifiers is significant, and an effort to use these answers more appropriately is a reasonable one.

Let us denote by $a1$ and $a2$ the top two choices (the first and the second respectively) of the structural feature classifier, and by $b1$ and $b2$ the top two choices of the statistical feature classifier. Some relations between the recognition outputs of the classifiers are given in Table 3.

EVENT	Recognition rate Test set (%)
$a1$ **and** $b1$	86.26
$a1$ **or** $b1$	96.86
$a1$ **or** $a2$ **or** $b1$ **or** $b2$	98.77

Table 3. Some relations between classifier decisions.

These observations confirm that the "top two" classifier outputs offer significant number of right answers, and an appropriate cooperation could lead to improvement of the recognition and reliability rates of the system.

4.2. Rejection criteria

To improve reliability of the system we will use rejection criteria as a part of the cooperation scheme. Let us denote by A_1, A_2 the best two outputs of the first and by B_1, B_2 the best two outputs of the second classifier (let us note that A_1, A_2, B_1, B_2 are classifier outputs, not class labels). Then, the simplest rejection criteria for the individual classifiers can be defined by

α	Structural features				Statistical features			
	Recog.%	**Miscl.%**	**Rejec. %**	**Reliab.%**	**Recog. %**	**Miscl.%**	**Rejec.%**	**Reliab.%**
0	90.34	9.66	0	90.34	92.77	7.23	0	92.77
$2x10^{-5}$	88.43	7.00	4.57	92.66	91.49	5.45	3.06	94.37
$4x10^{-5}$	86.24	5.25	8.51	94.26	89.77	4.31	5.92	95.42
$6x10^{-5}$	83.88	3.85	11.08	94.33	87.47	3.53	9.00	96.12
$8x10^{-5}$	80.86	2.99	16.15	96.43	79.77	4.09	16.14	95.12

Table 4. The classifier decisions using rejection criteria.

$$O1(\boldsymbol{x}) = \begin{cases} a1 = c(A_1), & if|A_1 - A_2| \geq \alpha \\ M+1, & otherwise \end{cases},$$

$$O2(\boldsymbol{x}) = \begin{cases} b1 = c(B_1), & if|B_1 - B_2| \geq \alpha \\ M+1, & otherwise \end{cases}$$

where c() is the function which gives the corresponding class for the classifier outputs, M is the number of classes and $M+1$ stays for the additional, rejection class. The parameter α, $0 \leq \alpha \leq 1$, is a threshold parameter. It controls the rejection rate according to the information for the certainty of the choice of the classes $a1$ and $b1$. The results of classifier outputs (Recognition, Misclassification, Rejection and Reliability = Recognition/(100%-Rejection)) using these rejection criteria are shown in Table 4.

Imposing stronger rejection criteria by increasing the value of the parameter α, the misclassification rate decreases while the rejection rate increases. The reliability of the system also increases until some specific value of the parameter α.

However, the results show that such rejection criterion is not a promising way to place a part of misclassified into the set of rejected patterns. Imposing a threshold level for accepting/refusing a given decision will not only discard a lot of misclassified digits, but will also discard a significant part of the well-recognized digits. These observations suggest that to achieve improved rejection criteria, it should be more useful to consider the activity values of all NN output nodes or to consider an integration of the activity values of two NN output nodes that “won the competition”. Considering the good recognition rates of the classifiers in cases of the “top two”outputs, we implemented a combination of both approaches.

We have investigated the values of the NN output nodes that give the first and the second decision label in each of the classifiers. In Table 5, we show the means $\mu_{1,k}$, $\mu_{2,k}$, $k = 1, 2$ and the standard deviations $\sigma_{1,k}$, $\sigma_{2,k}$, $k = 1, 2$ for the differences between the values of the NN output nodes that gave the first and the second decision label $(A_1 - A_2, B_1 - B_2)$ in cases of specific events.

Classification module	**Event**	**Mean** $(\mu_{i,k})$	**St. deviation** $(\sigma_{i,k})$
Structural Features	1^{st} decision is the right one	$\mu_{1,1} = 1.4908$	$\sigma_{1,1} = 0.5237$
	2^{nd} decision is the right one	$\mu_{1,2} = 0.0249$	$\sigma_{1,2} = 0.1455$
Statistical Features	1^{st} decision is the right one	$\mu_{2,1} = 1.6179$	$\sigma_{2,1} = 0.4669$
	2^{nd} decision is the right one	$\mu_{2,2} = 0.0258$	$\sigma_{2,2} = 0.1551$

Table 5. Some statistic values of classifier outputs.

This information, obtained on the trained classifiers, will be used in our rule-based cooperation schemes and in the corresponding rejection criteria incorporated in these schemes.

5. Multistage classification and final result

The main aim of the first decision fusion phase is to classify input patterns as much as possible, keeping a low misclassification rate. So, the result of this phase will be a set of correctly classified patterns and a set of "hard" patterns, rejected from this phase of classification. To complete this task we will use the information provided by the member classifiers in the following form:

— The "top two"class labels of each of the classifiers: $a1$ and $a2$ for the structural, and $b1$ and $b2$ for the statistical feature classifier;

— The differences between the "top two"classifier outputs $d_1 = A_1 - A_2$, $d_2 = B_1 - B_2$ and the intervals $[\mu_{\mathbf{i},k}, \mu_{\mathbf{i},k} + s \cdot \sigma_{\mathbf{i},k}]$, $i, k = 1, 2$, where these values belong. The parameter s is a suitable constant.

These values allow introduction of an additional certainty measure for the reliability of the classifier outputs. Thus, along with the first and the second decisions of the classifiers, we will consider the intervals where the differences between the corresponding classifier outputs belong (see Table 6).

The events $f1$, $f2$, $g1$ and $g2$ occur, when the difference between the "top two" classifier outputs falls near to the averages of the same differences for the corresponding classifier decisions $a1$, $a2$, $b1$ and $b2$. These averages are given in Table 5 and they are obtained on the trained classifiers. Thus, we have additional proof that the corresponding classifier decision $a1$, $a2$, $b1$ or $b2$ is correct.

Using the above information, we introduced the rule-based decision procedure for this phase of the classification process. We investigated some rule-based decision schemes in order to obtain a low misclassification rate with as high as possible recognition rate. In Table 7, some of the examined rule-based scheme with low misclassification rates and the corresponding results are presented.

There is no a general guideline how to chose the value of the parameter s, as well as, how to design the "best" rule-based decision scheme. These choices of the parameter s and the rule-based decision scheme follow the idea of minimizing the classifier misclassification rate.

Our experiments show that the lowest misclassification rate is provided by the rule 1 in Table 7. According to this rule, if the member classifiers give the same class as the first choice, we take that decision as final. Otherwise, we take the second choice label of better individual classifier (in our case the statistical feature classifier) $b2$ only if it gives the same decision label as the first choice of the structural feature classifier $(b2 = a1)$, and both classifiers are enough sure in their decisi ons $(g2$ **and** $f1)$. Finally, we take the second choice $a2$ of the structural feature classifier if it gives the same decision label as the first choice of the statistical feature classifier $(a2 = b1)$, and both classifiers are enough sure in their decisions $(f2$ **and** $g1)$.

Using a strength rule-based cooperation in the first phase of the classification, we minimized the misclassification rate of our cooperation scheme. In the second phase, we used a statistical cooperation scheme on the reduced set of "hard" patterns that were rejected during the rule-based classification. For this phase of clas-

Classification module	**Notation**
Structural Features	$f1$ **if** $d_1 \in [\mu_{1,1} - s \cdot \sigma_{1,1}, \mu_{1,1} + s \cdot \sigma_{1,1}]$
	$f2$ **if** $d_1 \in [\mu_{1,2} - s \cdot \sigma_{1,2}, \mu_{1,2} + s \cdot \sigma_{1,2}]$
Statistical Features	$g1$ **if** $d_2 \in [\mu_{2,1} - s \cdot \sigma_{2,1}, \mu_{2,1} + s \cdot \sigma_{2,1}]$
	$g2$ **if** $d_2 \in [\mu_{2,2} - s \cdot \sigma_{2,2}, \mu_{2,2} + s \cdot \sigma_{2,2}]$

Table 6. Notation of the characteristic events.

#	Rule	s	Recog. %	Misc. %	Rejec. %	Reliab. %
1.	**if** $a1 = b1$ **then** $c \leftarrow b1$ **elseif** $(g2$ **and** $f1)$ **and** $b2 = a1$ **then** $c \leftarrow b2$ **elseif** $(f2$ **and** $g1)$ **and** $a2 = b1$ **then** $c \leftarrow a2$ **else** REJECT	$s = 0.2$	86.33	1.5	12.18	98.30
2.	**if** $g1$ **or** $a1 = b1$ $c \leftarrow b1$ **elseif** $f1$ **then** $c \leftarrow a1$ **elseif** $(g2$ **or** $f1)$ **and** $b2 = a1$ **then** $c \leftarrow b2$ **elseif** $(f2$ **or** $g1)$ **and** $a2 = b1$ **then** $c \leftarrow a2$ **else** REJECT	$s = 2$ $s = 0.2$	94.13 88.26	4.76 2.05	1.11 9.69	95.19 97.73
3.	**if** $g1$ **and** $a1 = b1$ $c \leftarrow b1$ **elseif** $f1$ **then** $c \leftarrow a1$ **elseif** $g2$ **and** $b2 = a1$ **then** $c \leftarrow b2$ **elseif** $f2$ **and** $a2 = b1$ **then** $c \leftarrow a2$ **else** REJECT	$s = 2$ $s = 1$	91.19 85.97	5.21 2.24	3.60 11.79	97.46 94.78
4.	**if** $g1$ **or** $a1 = b1$ $c \leftarrow b1$ **elseif** $f1$ **then** $c \leftarrow a1$ **elseif** $g2$ **or** $b2 = a1$ **then** $c \leftarrow b2$ **elseif** $f2$ **or** $a2 = b1$ **then** $c \leftarrow a2$ **else** REJECT	$s = 2$ $s = 1$	94.30 93.90	5.70 5.01	0.0 1.09	94.30 94.93

Table 7. Various rule-based strategies and the corresponding recognition results.

sification we examined some statistical cooperation schemes. Best results were provided by the generalized committee cooperation scheme (Bishop 1995), described in section 3.

Thus, in the second decision fusion phase, we used the generalized committee cooperation scheme to classify rejected digit patterns. This combination of two different decision fusion schemes resulted in a high reliability recognition system.

In Table 8, the recognition results of the final classification are presented.

The results of the final classification show that the primary goal of the research is achieved. We have obtained a high reliability rate for the recognition system that also keeps a high recognition rate.

6. Conclusion

In this paper, we discussed the high reliability system for hand-written digit recognition. We have shown a possibility for cooperation of classifiers based on different feature families using

System	Recog.%	Misc.%	Rejec. %	Reliab.%
Rule 1 (Table 7) + generalized committee	95.10	1.64	3.26	98.30

Table 8. The final classification.

committee classifiers. For each of the feature families a simple MLP NN has been designed. We examined different cooperation schemes. The results show that the non-linear cooperation schemes give better recognition results.

To design high reliability recognition system, we introduced a multistage cooperation scheme with a rejection criterion. In the first decision fusion phase, we used rule-based cooperation scheme in order to provide as high as possible recognition rate, keeping misclassification rate as low as possible. In the second decision fusion phase, we used the generalized committee cooperation scheme only on the set of rejected pattern. The results of the final classification show that we proposed a successful approach for reliability control in committee classifier environment. Additionally, the results indicate that a suitable cooperation of statistical and rule-based decision schemes is a promising approach in handwritten recognition systems.

References

[1] Y.L. BARABAS, *Kollektivnie statisticeskie resenia pri raspoznavanii,* Moskva, Radio i svjaz, (*Collective statistical solutions for recognition,* in Russian), 1983.

[2] C.M. BISHOP, *Neural networks for pattern recognition,* Clarendon Press, Oxford, 1995.

[3] G. BUREL, I. POTTIER, J.Y. CATROS, Recognition of handwritten digits by image processing and neural network, *Proc. of the int. joint conference on neural networks,* IJCNN 3, (1992) pp. 666–671, Beijing.

[4] D. CAKMAKOV, Curve Matching Using Turning Functions", *Proc. of the Int. conf. on signal and image processing, SIP'98,* (1998) pp. 588–592, Las Vegas, USA.

[5] T. CIBAS, F.F. SOULIE, P. GALLINARI, S. RAUDYS, Variable selection with optimal cell damage, *Proc. of the Int. conference on artificial neural networks, ICANN'94,* (M. Marinaro and P.G. Morasso Eds.), Vol. 1, (1994) pp. 727–730, Springer-Verlag.

[6] J. DAHMEN, D. KEYSERS, H. NEY, Combined Classification of Handwritten Digits Using the 'Virtual Test Sample Method', *MCS 2001, LNCS 2096,* (J. Kittler and F. Roli Eds.), (2001) pp. 109–118, Springer-Verlag.

[7] B. DUERR, W. HAETTICH, H. TROPF, G. WINKLER, A combination of statistical and syntactical pattern recognition applied to classification of unconstrained handwritten numerals, *Pattern Recognition,* 12 (1980), 189–199.

[8] R.P.W. DUIN, D.M.J. TAX, Experiments with Classifier Combining Rules, *MCS 2000, LNCS 1857,* (J. Kittler and F. Roli Eds.), (2000) pp. 16–29, Springer-Verlag.

[9] L. HEUTTE, J.V. MOREAU, T. PAQUET, Y. LECOURTIER, C. OLIVIER, Combining structural and statistical features for the recognition of handwritten characters, *Proc. of the 13th Int. conf. on pattern recognition,* (1996) B74.4, Vienna, Austria,

[10] T.K. HO, J.J. HULL, S.N. SRIHARI, Decision combination in multiple classifier systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 16(1) (1994), 66–77.

[11] Y.S. HUANG, C.Y. SUEN, An Optimal Method of Combining Multiple Classifiers for Unconstrained Hand-written Numeral Recognition, *Proc. of IWFHR-3,* (1993) 11–20, Buffalo NY.

[12] F. KIMURA, M. SHRIDAR, Handwritten numerical recognition based on multiple algorithms. *Pattern Recognition,* 24(10) (1991), 969–983.

[13] J. KITTLER, Combining classifiers: A theoretical framework, *Pattern Analysis and Applications,* 1(1) (1998), 18–27.

[14] J. KITTLER, M. HATEF, R.P.W. DUIN, J. MATAS, On Combining Classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20 (3) (1998), 226–239.

[15] Y. LECUN, L.D. JACKEL, L. BOTTOU, A. BRUNOT, C. CORTES, J.S. DENKER, H. DRUCKER, I. GUYON, U.A. MULLER, E. SACKINGER, P. SIMARD, V. VAPNIK, Comparison of Learning Algorithms for Handwritten Digit Recognition, *International Conference on Artificial Neural Networks,* (F. Fogelman and P. Gallinari Eds), (1995) pp. 53–60, Paris.

[16] H. LIU, H. MOTODA, *Feature Selection for Knowledge Discovery and Data Mining,* Kluwer Academic Publisher, Boston, 1998.

[17] Z. LOU, K. LIU, J.Y. YANG, C.Y. SUEN, Rejection criteria and pairwise discrimination of handwritten numerals based on structural features, *Pattern Analysis and Applications,* 2(3) (1999), 228–238.

[18] V. RADEVSKI, Y. BENNANI, Reliability control in committee classifier environmen, *Int. joint conference on neural networks, IJCNN 11,* Vol. III, (2000) pp. 561–565. Como, Italy.

[19] F. ROLI, G. GIACINTO, G. VERNAZZA, Methods for Designing Multiple Classifier Systems, *MCS 2001, LNCS 2096,* (Kittler and F. Roli Eds.), (2001) pp. 78–87, Springer-Verlag.

[20] T. SUZUKI, H. NISHIDA, Y. NAKAJIMA, H. YAMAGATA, M. TACHIKAWA, G. SATO, A Handwritten Character Recognition System by Efficient Combination of Multiple Classifiers. *Int. association for pattern recognition workshop on document analysis systems,* (1995), pp. 169–187, World Scientific, Singapore.

[21] K. TUMER, J. GHOSH, Linear and order statistics combiners for pattern classification, *A. Combining Artificial Neural Nets,* (Sharkey Ed.), (1999) pp. 127–162, Springer-Verlag.

[22] L. XU, A. KRZYZAK, C.Y. SUEN, Methods of combining multiple classifiers and their applications to handwritten recognition. *IEEE Transaactions on Systems, Man and Cybernetics,* 22(3) (1992), 418–435.

[23] S. YAMAGUCHI, T. TSUTSUMIDA, F. KIMURA, A. IWATA, Study on Multi-Expert Systems for Hand-printed Numeral Recognition, *Progress in Handwriting Recognition,* (A. C. Downton – S. Impedovo Eds.) (1997), pp. 285–292, World Scientific Publ., Singapore. 11

Received: December, 2001
Accepted: September, 2002

Contact address:
Dusan Cakmakov
University "Sv. Kiril i Metodij"
Faculty of Mechanical Engineering
Department of Mathematics and Computer Science,
Karpos II bb, POBox 464
1000 Skopje, Macedonia
e-mail: dusan@mf.ukim.edu.mk

Vladimir Radevski
University Galatasaray
Faculty of Engineering and Technology
Department of Computer Engineering
Ciragan Cad. 102, 80840 Ortakoy
Istanbul, Turkey
e-mail: radevski@netone.com.tr

Younes Bennani
University Paris 13
LIPN-CNRS
Avenue J.B. Clement
93430 Villetaneuse, France
e-mail: younes@lipn.univ-paris13.fr

Dejan Gorgevik
University "Sv. Kiril i Metodij"
Faculty of Electrical Engineering
Department of Computer and Information Technology
Karpos II bb, POBox 574
1000 Skopje, Macedonia
e-mail: dejan@etf.ukim.edu.mk

DR. DUSAN CAKMAKOV obtained his Ph.D. in computer science from the faculty of Electrical Eng. & Computer Science, University "St. Cyril and Methodius", Skopje, Macedonia in 1992. In 1994 he was research scientist on Hypermedia and Information Retrieval at the University of Southampton, UK, in 1996 visiting professor at the University Paris XIII, France and in 2001 visiting fellow at the Oxford University, UK. Currently, Dr. Cakmakov is an Associate Professor in computer science and applied mathematics at the Faculty of Mechanical Eng., University "St. Cyril and Methodius", Skopje, Macedonia. His current research interests include pattern recognition and content-based information retrieval. He is author of more then 30 papers in refereed conference proceedings, journals and contributions in books.

DR. VLADIMIR RADEVSKI obtained his M.S. degree form the Institute of Mathematics, Faculty of Natural Sciences, University of Belgrade, Yugoslavia in 1994 and his Ph.D. in computer science from the Institute Galilee, University of Paris 13, France in 2000. Currently, Dr. Radevski is an Assistant Professor in computer science at the Faculty of Engineering and Technology, University Galatasaray, Istanbul, Turkey. His current research interests include pattern recognition and computer vision. He is author of a number of papers in pattern recognition area, published in refereed conference proceedings and journals.

DR. YOUNES BENNANI received his B.S. in mathematics and computer science from Rouen University, and his M.S. and Ph.D. degrees in computer science from the University of Paris 11 (ORSAY) in 1992, and HDR (Habilitation a Diriger des Recherches) degrees in computer science from the University of Paris 13 in 1998. He is currently Professor at the University of Paris 13, with research interest in theory of neural networks, architecture and algorithm design and analysis, statistical pattern recognition, datamining. He is also interested in the application of these models to speech recognition, speaker identification, languages identification, diagnosis of complex systems, abnormal situation detection and webmining. He has published more than 50 papers in refereed conference proceedings, journals or as contributions in books.

DEJAN GORGEVIK received his M.S. degree in computer engineering from the Faculty of Electrical Eng. & Computer Science, University "St. Cyril and Methodius", Skopje, Macedonia in 1997. Currently, Mr. Gorgevik is a Senior Teaching and Research Associate at the Faculty of Electrical Eng. & Computer Science, University "St. Cyril and Methodius", Skopje, Macedonia. He is working on his Ph.D. in the area of combining classifiers for handwritten character recognition. The scope of his interests encompasses: pattern recognition, image processing, computer graphics, and computer vision. He is author of a number of papers published in refereed conference proceedings.