HANDWRITTEN DIGIT RECOGNITION BY COMBINING SUPPORT VECTOR MACHINES USING RULE-BASED REASONING

Dejan Gorgevik, Dusan Cakmakov[*], Vladimir Radevski[**]

University "Sv. Kiril i Metodij", Faculty of Electrical Eng., Department of Computer and Information Technology, Karpos II bb, POBox 574, 1000 Skopje, Macedonia
dejan@cerera.etf.ukim.edu.mk

[*]University "Sv. Kiril i Metodij", Faculty of Mechanical Eng., Department of Mathematics and Computer Science, Karpos II bb, POBox 464, 1000 Skopje, Macedonia
dusan@ereb.mf.ukim.edu.mk

[**]University Galatasaray, Faculty of Engineering and Technology, Department of Computer Engineering, Ciragan Cad. 102 Ortakoy, Istanbul, Turkey
radevski@netone.com.tr

Abstract: *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. In this paper, the cooperation of two feature families for handwritten digit recognition using SVM (Support Vector Machine) classifiers will be examined. We investigate the advantages and weaknesses of various decision fusion schemes using rule-based reasoning. The obtained results show that it is difficult to exceed the recognition rate of the classifier applied straightforwardly on the feature families as one set. However, the rule-based cooperation schemes enable an easy and efficient implementation of various rejection criteria that leads to high reliability recognition systems.*

Keywords: structural, statistical, features, decision fusion, rejection, reliability

1. Introduction

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 many pattern recognition applications. Data from more than one source that are processed separately 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, 1988; Tumer, 1999).

In this paper we discuss classification systems for handwritten digit recognition using two different feature families and SVM classifiers (Burges, 1998). Our feature families are referenced as structural and statistical feature sets (Cakmakov, 1999; Radevski, 2000), and they differ (especially structural features) from the feature sets with the same reference used in other systems for handwritten character recognition (Duerr, 1980; Heutte, 1996). We start with a SVM classifier applied on both feature families as one set. These results serve as a basis for future investigations. Further, we used two SVM classifiers that work on the different feature families for the same digit image. As the feature sets "see" the same digit image from two different points of view, we examined the possibility of decision fusion using rule-based

reasoning. Different rule-based cooperation schemes are examined and corresponding recognition results are presented. In order to improve the system reliability, we introduced rejection criteria that are "natural" part of rule-based cooperation.

The presented results show that it is difficult to achieve the recognition rate of a single classifier applied on the feature set that includes both feature families by rule-based reasoning applied on the individual classifier decisions. However, rule-based decision fusion enables implementation of various rejection criteria that leads to high reliability recognition systems. Additionally, separate classifiers designed for separate feature families reduce classifier complexity and offer better possibilities to understand the role of the features in the recognition process.

2. The system architecture

The recognition system is constructed around a modular architecture of feature extraction and digit classification units. Preprocessed image is an input for the feature extraction module, which transfers the extracted features toward SVM classifiers (Fig. 1).
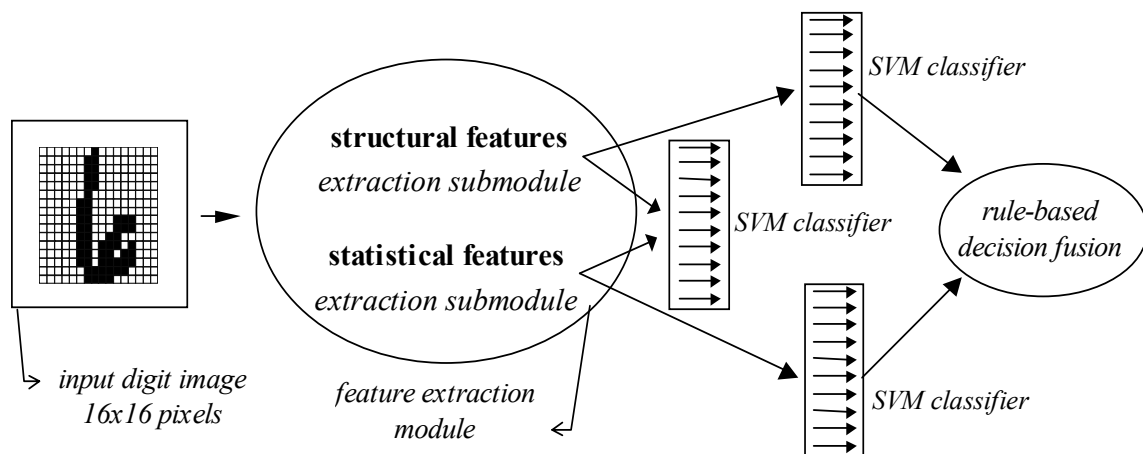

Fig. 1: The system architecture

From the digit images with resolution of 128×128 pixels, we have obtained 16×16 binary images on which the smoothing and centralizing preprocessing techniques have been applied. We have extracted 116 features that are classified as 54 structural and 62 statistical. The both feature families as one set are forwarded to the SVM classifier and obtained results are basis for future comparisons.

The structural and statistical feature sets are also forwarded to the separate SVM classifiers, and obtained results are combined using rule-based reasoning. On this level, rejection criteria are introduced and the corresponding system reliabilities are calculated.

3. The Handwritten Digit Data Base and Feature Extraction

The database for our experiments is an extraction of the NIST (National Institute of Standards and Technology) segmented handwritten digit database. The digit images are in 128×128 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. 2, a fragment and the composition of the digit database are presented.

	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. 2: A fragment and composition of the digit database

To create the structural feature set we define a reasonable set of elementary shape primitives for digit constructions. We have proposed 27 elementary primitives showed in Fig. 3. 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 primitives is 54, and that is the number of the elements in the structural feature set.

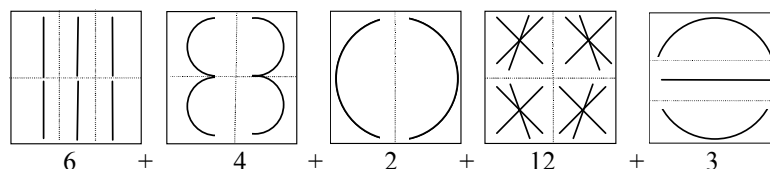

| 6 | + | 4 | + | 2 | + | 12 | + | 3 |

Fig. 3: Image sub-regions and elementary primitives

The existing shape in each of those sub-regions is compared with the referent, idealized primitives in the same sub-regions whose existence is expected. The similarity measure between the found shape and the primitive is based on differences of changes of angles along both shapes, normalized to take values between 0 and 1. This similarity measure is a simplified variation of the curve matching technique described in (Cakmakov, 1998).

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

A digit image   Projection histograms 1-54
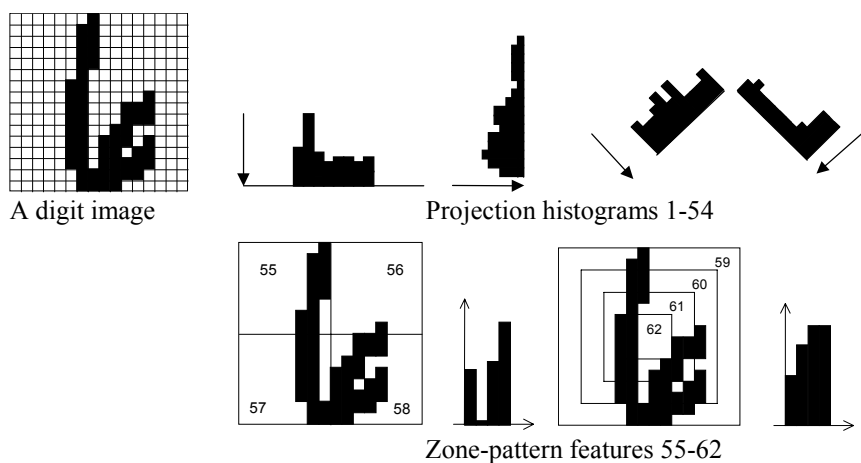
Zone-pattern features 55-62

Fig. 4: Projection histograms and zone-pattern features

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.

4. The Recognition Result

In the following table, the classification results along with the 99% confidence interval using both feature families as a feature set are presented. The classifier is SVM with linear and one nonlinear kernel, in our case Gausian.

Table 1: Recognition rates on the test set of SVMs using different kernel functions

Structural + Statistical features	Recognition rate - Test Set
Linear kernel	95.06% [94.33%, 95.78%]
Gausian kernel (σ=2.0)	97.53% [97.00%, 98.06%]

Considering Table 1, it is evident that SVM with Gausian kernel provides noticeably better recognition rate. Actually, our experiments showed that Gausian kernel provides better recognition rate then polynomial or sigmoidal kernel. Because of the large number of samples we have used SVMTorch, that is a more robust variation of SVM training software (Collobert, 2000).

5. The Rule-Based Decision Fusion

The rule-based decision schemes are designed around two SVM classifiers with Gaussian kernels. They perform the classification task on the sets of structural and statistical features respectively.

Let us denote by *a1, a2* and *a3* the first, the second and the third choice of the structural feature classifier, and by *b1, b2* and *b3* the first, the second and the third choice of the statistical feature classifier for a given pattern. The recognition rate of the classifiers for structural and statistical feature sets together with "top two" and "top three" cases (correct decision is among the first two and the first three classifier choices) are evaluated and given in Table 2.

Table 2: Recognition rates of the individual classifiers (test set)

Classification module	Recognition rate	"Top two" case	"Top three" case
structural features	95.12 %	98.22%	99.08%
statistical features	96.97 %	98.99%	99.45%

The results show that the statistical feature set has stronger discrimination power and provide better recognition rate. However, the recognition rate of the statistical feature set is more then 0.5 percent lower then the recognition rate of the classifier applied to the complete feature set (see Table 1). Our experiments showed that the inclusion of additional choices (after the third) provides insignificant recognition rate improvement. Thus, in our cooperation schemes we have used only the top three choices of both classifiers.

The percentage of some characteristic events of classifier behaviors: "top one – top one", "top one – top two", "top two – top one", "top two – top two" etc. are evaluated and given in Table 3. These results give the upper limits of the recognition rates of any combination that utilizes corresponding individual classifier decisions. It is clear that the number of correct answers among the first choices of the classifiers is high.

Table 3: Some relations between classifier decisions

Correct decision	Number of samples (out of 5946)	Percentage
in {*a1, b1*}	5859	98.54%
in {*a1, b1, b2*}	5908	99.36%
in {*a1, a2, b1*}	5896	99.16%
in {*a1, a2, b1, b2*}	5919	99.55%
in {*a1, a2, a3, b1, b2, b3*}	5932	99.76%

Considering the results from Table 2 and Table 3, we can notice that feature families offer some level of complementary information about handwritten digits and fusion of the individual classifier decisions could result in a high reliability recognition system that will also keep a high recognition rate.

To improve reliability of the system we used rejection criteria as a part of the rule-based cooperation schemes. The results of classifier outputs (Recognition, Misclassification, Rejection and Reliability = Recognition/(100% −Rejection)) based on various rule-based cooperation schemes are evaluated and given in Table 4.

Table 4: Various rule-based strategies and corresponding recognition rates

#	Rule-based strategies	Recog.	Miscl.	Rejec.	Reliab.
1.	if *a1=b1* then .... *c=a1* else REJECT	93.79% (5577)	0.89% (53)	5.31% (316)	**99.06%**
2.	if *a1=b1* or *a1=b2* then .... *c=a1* else REJECT	94.75% (5634)	2.44% (145)	2.81% (167)	97.49%
3.	if *b1=a1* or *b1=a2* then .... *c=b1* else REJECT	**96.27%** (5724)	1.63% (97)	2.10% (125)	**98.33%**
4.	if *a1=b1* or *a1=b2* then .... *c=a1* elseif *b1=a2* then .... *c=b1* else REJECT	96.22% (5721)	2.62% (156)	1.16% (69)	97.35%
5.	if *a1=b1* or *a1=b2* or *a1=b3* then .... *c=a1* else REJECT	94.99% (5648)	3.26% (194)	1.75% (104)	96.68%
6.	if *b1=a1* or *b1=a2* or *b1=a3* then .... *c=b1* else REJECT	**96.79%** (5755)	1.95% (116)	1.26% (75)	**98.02%**
7.	if *a1=b1* or *a1=b2* or *a1=b3* then .... *c=a1* elseif *b1=a2* or *b1=a3* then .... *c=b1* else REJECT	96.22% (5721)	3.38% (201)	0.40% (24)	96.61%
8.	if *b1=a1* or *b1=a2* or *b1=a3* then .... *c=b1* elseif *a1=b2* or *a1=b3* then .... *c=a1* else REJECT	**97.26%** (5783)	2.34% (139)	0.40% (24)	97.65%

Four results in Table 4 deserve attention. Best reliability is obtained by the rule 1 (consensus) but the recognition rate is relatively weak. A good compromise is provided by rules 3 and 6, where we choose the first decision *b1* of the statistical feature classifier as a final decision *c*, if it is among the "top two" decisions (*a1, a2*) in the rule 3 and among the top three decisions (*a1, a2, a3*) in the rule 6 of the structural feature classifier. It seems that in this case the structural feature classifier gives a safety rule for the right decision. The reliabilities of 98.33% and 98.02% by recognition rates of 96.27% and 96.79% are noticeable results, better then some previous attempts using the same feature sets (Cakmakov, 1999; Radevski, 2000). The main contribution in this result is provided by the SVM classifier that showed superior generalization ability comparing to traditional MLP (Multi-Layer Perceptron) neural network.

On the other hand, best recognition rate is provided by the relatively complex rule 8. Unfortunately, this rule produces high misclassification rate that results in lower reliability. Let us notice that the recognition rate achieved by rule-based cooperation schemes is still about 0.25% lower then the recognition rate of the SVM that uses both feature families as one feature set (Table 1). It is probably due to greater "roughness" of our rule-based cooperation schemes that cannot be fine-tuned like most of the statistical cooperation schemes. However, the strength of rule-based reasoning usually provides improved system reliability.

There is no general guideline on how to chose the "best" rule-based scheme based on individual classifier decisions. However, it is possible to evaluate all "promising" rule-based schemes based on "top few" decisions, even in cases of more then two separate classifiers.

6. Conclusion

In this paper, we address some issues in designing high reliability system for hand-written digit recognition using SVM classifiers. We used two different feature families referenced as structural and statistical features. Decision level fusion is performed using rule-based reasoning. To examine possibilities for improving of the system reliability, we introduced rejection criteria in decision fusion schemes.

The presented results show that it is difficult to achieve the recognition rate of the single classifier applied on the feature set that includes both feature families by rule-based reasoning applied on the individual classifier decisions. However, the strength rule-based cooperation schemes enable an easy implementation of various rejection criteria that leads to high reliability recognition systems.

7. References

1. Burges C,. (1998), "A Tutorial on Support Vector Machines for Pattern Recognition", *Knowledge Discovery and Data Mining*, Vol. 2, pp. 1-47.

2. Cakmakov D., (1998), "Curve Matching Using Turning Functions", *Proc. of the Int. Conf. on Signal and Image Processing SIP'98*, Las Vegas, USA, pp. 588-592.

3. Cakmakov D. and Radevski V., (1999), "Experiments in NN based OCR using committee classifiers", *Proc. of 21st Int. Conf. Information Technology Interfaces, ITI'99*, Pula, Croatia, 1999, pp. 57-61.

4. Collobert R. and Bengio S., (2000), "Support Vector Machines for Large-Scale Regression Problems", IDIAP-RR-00-17, Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP), Case postale 609, CH-1920 Martigny, Switzerland. (www.idiap.ch/learning/SVMTorch.html)

5. Duerr B., Haettich W., Tropf H., and Winkler G., (1980), "A combination of statistical and syntactical pattern recognition applied to classification of unconstrained handwritten numerals", *Pattern Recognition*, Vol. 12, pp. 189-199.

6. Heutte L., Moreau J. V., Paquet T., Lecourtier Y., and Olivier C., (1996), "Combining structural and statistical features for the recognition of handwritten characters", *Proc. of the 13th Int. Conf. on Pattern Recognition*, 1996, pp. B74.4.

7. Kittler J., (1988), "Combining classifiers: A theoretical framework", *Pattern Analysis and Applications*, Vol. 1, pp. 18-27.

8. Radevski V. and Bennani Y. (2000), "Reliability control in committee classifier environment", *Int. joint conference on neural networks, IJCNN 11*, Como, Italy, Vol. III, 2000, pp. 561-565.

9. Tumer K., Ghosh J., (1999), "Linear and Order Statistics Combiners for Pattern Classification", in *Combining Articial Neural Nets*, A. Sharkey (Ed.), Springer-Verlag, 1999, pp. 127-162.