Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Pattern Recognition Letters 26 (2005) 369–379
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Analysis of errors of handwritten digits made by a multitude of classifiers

Ching Y. Suen *, Jinna Tan

Centre for Pattern Recognition and Machine Intelligence, Concordia University, CENPARMI, Suite GM 606, 1455 de Maisonneuve Blvd. West, Montréal, Québec, Canada H3G 1M8

Received 5 July 2004; received in revised form 18 October 2004
Available online 19 December 2004

Abstract

In this paper we describe an in-depth study on some data misclassified by a collection of classifiers produced by different authors. First of all, we divide the errors into three categories based on their quality and analyze their distributions according to category. Common errors made by three or more classifiers out of five have been identified and analyzed to deduce the reasons of misclassification. Finally, based on systematic analyses, two possible solutions to reduce errors and improve system reliability are proposed: (a) a verification module, and (b) combination of complementary multiple classifiers.

Keywords: Error analysis; Handwriting digit recognition; Multiple classifiers; Verifier

In memoriam

I was formally introduced to Prof. Azriel Rosenfeld by my host Prof. Murray Eden (Head of the Cognitive Information Processing Group of MIT) while I was visiting Cognitive Information Processing Group in the 70s. I met Prof. Rosenfeld many times at ICPRs (including the first one held in Washington, DC in 1973) and at numerous other occasions. Prof. Rosenfeld's total devotion to research and education and his immense outputs are always admired by our colleagues and students. Here at the Centre for Pattern Recognition and Machine Intelligence at Concordia University, we keep many of Dr. Rosenfeld's publications on image modeling and picture processing, and we are always fascinated by the ideas behind them. The last time I talked to him took place at ICPR-2002 in Quebec City after listening to the speeches

* Corresponding author.
E-mail addresses: suen@cenparmi.concordia.ca (C.Y. Suen), jnmtan@cenparmi.concordia.ca (J. Tan).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.10.019

given by him and other K.S. Fu Prize Winners. There Dr. Rosenfeld gladly accepted our invitation (Patrick Wang and me) to attend ICPR-2006 in Hong Kong.

In the past few years, we have given serious thoughts to strategic research in reducing errors in handwriting recognition, and this paper summarizes our recent study which forms the principal foundation towards the realization of an error free recognition system. We dedicate this paper to the memory of Prof. Rosenfeld.

(Ching Y. Suen)

1. Introduction

Recognition of unconstrained isolated handwritten numerals is an important area of OCR. It has applications in numerous practical environments including automatic mail sorting, bank cheque recognition, financial slip and business form processing, and so on. Many techniques have been developed to improve the recognition performance; some of the most popular techniques used include various feature selection schemes, classification methods, and different system architectures such as multi-experts, and verification modules. Some current methods have achieved a high recognition rate of more than 99%, such as the classifiers developed by Kussul and Baidyk (2003) and Liu et al. (2004), which achieve recognition rates of 99.51% on MNIST database and 99.15% on CENPARMI database, respectively. However, the problem of misrecognition persists despite intensive research efforts in combination of different feature sets and development of superb classifiers.

In financial applications, errors are less tolerable than rejections since much extra effort is required to detect and correct the errors; therefore, very high reliability is desired. Before the development of a system to reduce errors and achieve high reliability, we first need to study the misclassified data, understand the reasons of misclassification, and find ways of preventing their occurrence. Such study is helpful and advantageous in the research and development of systems with minimum errors. For the purpose of satisfying the requirement of high reliability, the classifiers must perform with minimal errors, or eventually be free from errors. This paper reports our in-depth study of the characteristics of misrecognized data aimed to deduce the reasons of misrecognition and probe the probability of avoiding these errors.

2. Database

In this analysis, misrecognized samples have been assembled from numerous databases to cover the best known data. They include errors made by 8 research teams on MNIST, CENPARMI, USPS, and NIST SD 19.

MNIST database is a widely used handwritten digit recognition benchmark. It contains binary images of handwritten digits constructed from the NIST's Special Database 3 and Special Database 1. SD-3 was collected from Census Bureau employees, while SD-1 was collected from high school students. The training set and the test set contains 60,000 and 10,000 images, respectively, half of them from SD-1 and the other half from SD-3. Images were first size normalized to fit into a box of 20×20 pixels, and then centered in a 28×28 image (LeCun et al., 1998).

The widely used CENPARMI database contains a training set of 4000 binary image samples, and a test set of 2000. These data were compiled from US zip codes of dead envelopes by researchers in CENPARMI (Ahmed and Suen, 1984). The images are of various sizes.

The US Postal Service (USPS) database contains 9289 handwritten digits, 7291 for the training set and 2007 for the test set. They were collected from envelopes in Buffalo. It is known that USPS test set is rather difficult. The human error rate is 2.5% (Bromley and Sackinger, 1991). It has also been shown that there are many confusing samples and obviously mislabeled data in the test set (Dong et al., 2001).

The NIST Special Database 19 (SD 19) is composed of 3669 full binary images of Handwritten Sample Forms. A total of 814,255 handwritten isolated characters have been segmented from these forms. The images vary considerably in size.

3. Misclassified data

Eight classifiers have been chosen in our study. They are typical and popular classifiers which operate under different classification approaches and are tested with different databases. Their performances are listed in Table 1.

GPR (Tan, 2004) (General-Purpose Recognizer) is a combination of multiple classifiers aimed at taking the most advantage out of the classification power of three classifiers. It was tested on MNIST database. VSVM[b] (Dong, 2003), VSV2 (Decoste and Scholkopf, 2002), and VSVM (Dong et al., 2002), are virtual support vector machines with different methods of generating support vectors. The first two were conducted on the MNIST database while the last one was tested on both CENPARMI and USPS database. LetNet-5 (LeCun et al., 1998) is a convolutional neural network comprising seven layers, including convolutional layers, sub-sampling layers and fully-connected layers. The experiment was conducted on MNIST database. The study of hierarchical Product of Experts (POE) (Mayraz and Hinton, 2002) was also carried out on MNIST database. Lastly, the ensemble of multi-layer perceptron MLP (Oliveira et al., 2003) was trained with the gradient descent applied to a sum-of-squares error function and tested on the NIST SD 19 database. The test set was extracted from hsf_7.

Table 1
Performance of classifiers

Classifier	Number of test samples	Recognition (%)	Error (%)
MNIST database			
GPR	10,000	99.06	0.94
VSVM[b]	10,000	99.62	0.38
VSV2	10,000	99.44	0.56
LeNet-5	10,000	99.18	0.82
POE	10,000	98.32	1.68
CENPARMI database			
VSVM	2000	98.7	1.3
USPS database			
VSVM	2007	97.66	2.34
NIST SD19 database			
MLP	30,000	99.16	0.84

The analysis was carried out on the entire set of samples misrecognized by all classifiers except MLP. MLP generated a large set of misrecognized samples due to the large test set. Only 119 of misrecognized samples are used in the analysis. The collection of misrecognized samples used in the analysis is presented in Appendix A.

4. Error analysis

4.1. Categories of misrecognized data

There are great shape variations in unconstrained handwritten numerals. A handwritten numeral can be so poorly formed that it resembles little its printed form. It may not look like any numeral or it may look like another numeral. These are the possible reasons of misclassification by human beings. In the collection of misrecognized data, some of them are clear and in a typical shape of a numeral that will not be misrecognized by human beings, while some of them are truly ambiguous or poorly degraded. The portion of misrecognized data which are easy for human beings should also be recognized by the classifiers. In this section, we will study how large portions of misrecognized data are easily recognized by human beings.

After careful inspections by five persons based on majority vote, we divided the collection of misrecognized data into three categories for further investigation from a human's perspective. The distributions of these three categories are listed in Table 2 for each classifier.

- Category 1 is for the images that are easily confused with other numerals because of the similarity of their primitives and structures. As we know, some pairs of numerals are more easily confused than others; such as 4–9, 0–6, and 3–5. Images in this category usually belong to these confusing pairs. Some of such images are shown in Fig. 1(a). Numeral 4 is confusing with numeral 9 since the curvature changes of the upper 'U' curve is smooth; the first numeral 6 is confusing with numeral 4 which is written in

Table 2
Three categories of misrecognized data

Classifier	Total number of errors	Category 1	Category 2	Category 3
MNIST database				
GPR	94	24	11	59
VSVM[b]	38	15	6	17
VSV2	56	15	9	32
LeNet5	82	17	14	51
POE	168	41	9	118
CENPARMI database				
VSVM	26	6	4	16
USPS database				
VSVM	47	13	13	21
NIST SD 19 database				
MLP	119	30	8	81
Sum	630	161	74	395
Percentage (%)	100	25.56	11.75	62.70

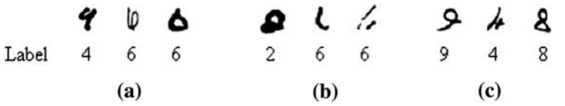


Fig. 1. Samples of misclassified data: (a) Category 1, (b) Category 2, and (c) Category 3.

one single stroke; the second numeral 6 is confusing with numeral 0 because the stroke on top of the circle is too short.

- Category 2 is for the images that humans have difficulty in identifying them because of noise, filled loop, cursive writing or over-segmentation, etc. Degradation and distortion are probably due to the quality of the scanner, the width of the tip of the writing instrument, size normalization, and the writing habit of certain people. Some samples belonging to Category 2 are presented in Fig. 1(b). The numeral 2 is degraded because of the thick tip of the pen: the upper 'D'' curve touches the bottom part of numeral two and the strokes of the bottom part overlap, which makes the image hard to be identified. The first numeral 6 is incomplete: some part of the six is missing; the bottom part of the second numeral 6 is broken, which is probably due to a faulty underline removal algorithm.
- Category 3 is for the images that are easily recognized by humans without any ambiguity. Images in this category are clear and unambiguous. They have all the necessary structural primitives, and have typical connectivity of the primitives. Fig. 1(c) presents three instances belonging to this category.

About one quarter of the misrecognized data belong to Category 1. It is normal that images in this category are misrecognized since their shapes are ambiguous. Images in Category 1 have similar structures but with local characteristics different from their confusing numerals. Some portions of images in Category 1 can be recognized correctly if their local differences are closely examined. However, since some of them can confuse even human beings, it is hard to say whether the original labels in the database or the estimated labels were correct. People have different opinions in identifying these images. Therefore, further studies need to be done in an attempt to recognize only parts of this category that can be identified consistently by most people. The rest should be rejected.

It is also reasonable that images belonging to Category 2 are misrecognized since the information obtained from the image is insufficient, incorrect, or misleading. For instance, it is hard to

Fig. 2. Samples of numeral one misrecognized as two.

identify a numeral six by seeing only the right half of it because it can be part of a six or a zero. It is extremely difficult to recognize this type of numerals, so the best way of handling them is to reject them. Fortunately, the quantity of images belonging to Category 2 is small.

There are 62.70% of the collection of misrecognized images which belong to Category 3. Humans have no difficulty in identifying them. Some of them are clear though they may not have the typical global structure of a numeral. The reason of misrecognition might be the lack of training samples that have these structures. For example, Fig. 2 presents some samples of numeral one that are clear and unambiguous, but all of them are misrecognized as numeral two. It is not difficult for human beings to identify them correctly. However, because this writing style is not typical and there are only a few samples written in this style in the training set, the classifiers do not have enough samples to learn from. Hence, these samples of numeral one are misclassified. On the other hand, some of these images are written clearly in the standard shapes of numerals; it is surprising to see that the classifiers misrecognized them. Images in this category form the targets of further investigation to improve the recognition performance.

Table 3
Distribution of common errors

No. of classifiers which made the same error	1	2	3	4	5	Sum
No. of errors	156	47	31	15	7	256
Percentage (%)	60.94	18.36	12.11	5.86	2.73	100

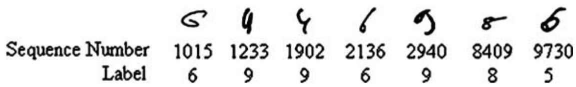


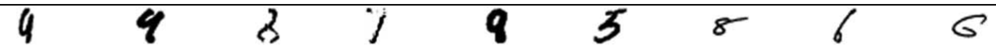
Fig. 3. Common errors of all classifiers.

4.2. Common errors of multiple classifiers

We also observed that some images are misrecognized by one classifier, some are misrecognized by more than one classifier, and some are misrecognized by the majority of classifiers. Since most of the classifiers conducted experiments on the MNIST database, we focus our analysis only on the common errors made by the five classifiers using this database: GPR, VSVM^b, VSV2, LeNet5 and POE, as summarized in Table 3. There are 256 errors made in MNIST database by the five classifiers (errors occur in more than one classifier is counted as one error). The majority of errors are made by one classifier. Only 2.73% errors occur in all the classifiers; 5.86% errors occur in four classifiers. The seven errors made by all classifiers are shown in Fig. 3.

Some images that are misrecognized by more than two classifiers are chosen and a thorough investigation for reasons of misclassification is

Table 4
Common mistakes of classifiers



Sequence no.	1233	2131	583	6577	948	2036	8409	2136	1015
Label	9	4	8	7	8	5	8	6	6
GPR	×	×			×		×	×	×
VSVM^b	×	×	×	×	×	×	×	×	×
VSV2	×	×	×	×	×	×	×	×	×
LeNet5	×	×	×			×	×	×	×
POE	×		×	×	×		×	×	×

presented in this section. The images of the errors and the corresponding classifiers which make them are shown in Table 4.

The first two images belong to Category 1. The first image is misidentified by all classifiers as numeral 4 while the second image is misrecognized by four classifiers as numeral 9. This result is not surprising since these two images look confusing enough to human beings. Numerals 4 and 9 form one of the most confusing pairs of numerals, even though they have salient differences in their printed form. However, when they are written by hand, particularly written cursively, their local differences become small and hardly detectable by human beings.

The 3rd to 5th images are perturbed numerals which belong to Category 2. The circle at the bottom of the first image is over-segmented and the right part of the upper circle disappeared due to the writing habit of the writer. The upper right portion of the numeral seven is probably lost in the scanning process or due to the writing instrument. The fifth image is a typical perturbation caused by the wide tip of the writing instrument, which filled the lower circle of the numeral eight. The image looks somewhat like a numeral nine. Some classifiers perform better in identifying these seriously perturbed numerals. The GPR is able to identify the first two images while LeNet-5 is able to recognize the last two.

The last four images are easily recognized by humans. They belong to Category 3. The images with sequence numbers of 2036 and 8409 have clear and typical structures of numerals eight and five, respectively. However, unsatisfactory results were produced by the selected classifiers. The numeral five can be recognized correctly only by classifiers GPR and POE. The numeral eight was misrecognized by all five classifiers. It is surprising that all classifiers misrecognized it as numeral five since there is nothing in common between this image and numeral five. All classifiers also misrecognized the two instances of numerals six whose sequence numbers are 2136 and 1015 respectively. As we can see, these two samples of numeral six do not have a typical shape of numeral six but still can be easily recognized by human beings with the structure of a circle at the bottom and a barely vertical stroke on top. Humans can identify the image 2136 as a six easily even though the circle is small. Similarly, although the image 1015 is slanted, it has the structure of a typical numeral six; therefore, it can be identified by human beings easily. However, all classifiers misrecognized these two numerals. The main reason may be due to the lack of training samples. Numeral six 2136 has a very small circle in the bottom part, which is unusual in numeral six; numeral six 1015 has a big slant. Thus, there are not enough samples that are similar to these two numerals to train the recognizer. Though there are a lot of training samples with a structure similar to the numeral eight 8409, we observed that the ratio of height to width of the minimum rectangle circumscribing the numeral eight 8409 is different from other instances of numeral eight. Usually, for a numeral eight, the ratio of height to width of the minimum rectangle is larger than 1. However, the ratio is smaller than 1 for the numeral eight 8409 because of the long stroke at the top, which is connected to the right part of the upper circle of the eight. In this respect, the numeral eight 8409 is also a rare shape indicating the lack of training samples for the recognizer.

4.3. Summary of reasons for misclassification

Based on the analyses presented in Sections 4.1 and 4.2, the reasons of misclassification can be summarized as follows:

- Confusing nature of some pairs of numerals: they have similar structures and shapes, e.g. numerals 4–9, 0–6.
- The width of the tip of the writing instrument. This causes some circles to be filled such as the lower circle of numeral eight; or some strokes to touch one another, e.g. one end of the top D curve of numeral three touches the other end of the D curve to form a circle.
- Low quality of the images due to the scanning process, writing instrument or size normalization. Some parts of the image are broken or some noises are generated; thereby erroneous features are grown and detected.

- Cursive writing. Some numerals are so cursive that some strokes become connected in unusual ways. Hence, incorrect structural features are detected.
- Over or under segmentation, or mistake in the process of removing underlines. They eliminate some parts of the numerals.
- Lack of training samples of some numerals that are written in atypical style.

4.4. Possible solution of reducing errors

According to the above analyses, one possible solution to reduce the error is the implementation of a verification module. The goal of a verifier is to evaluate precisely results produced by the classification stage in order to compensate for its weakness. This evaluation, consequently, should reduce the error rate and make the entire system more reliable. There are three types of verifiers as defined by Takahashi and Griffin (1993): absolute verification for each class (i.e., is it a numeral "1"?), pairwise verification between two classes (i.e., is it a numeral "4" or "9"?), and verification in clustered, visually similar, classes (i.e., is it a "0", "6", or "9").

Misclassified data of Category 1 can be further examined by pairwise verification, which is easier to concentrate on the local differences between two classes. In this type of verification, a set of pairwise verifiers is built according to the frequency of confusions among digits. When the confidence value of a pattern is lower than a threshold, the pairwise verifier is applied to verify whether the classification result is correct. For example, since numeral 4 is confusing with numeral 9 and 6, thus pairwise verifiers 4–9 and 4–6 are built. For a pattern that is classified as numeral 4, if the confidence value of the classification is low, the two pairwise verifiers are applied to check whether the pattern is more likely to be numeral 4 or its confusing numerals. On the other hand, misclassified data of Categories 2 and 3 can be verified by using absolute verification. The verification module can be built by identifying the structural primitives of a pattern and the geometric relations among them. The distorted numerals may not have all the necessary structural primitives and characteristics to be identified correctly, but neither of them have the characteristics of the wrongly classified numeral. By examining their primitives, the absolute verifier should be able to detect that the classification result is wrong since they do not have the necessary characteristics of the estimated numeral. In this way, errors can be detected and reduced by the absolute verifier. This also applies to the misclassified data of Category 3. These easy-to-recognize images have clear and unambiguous structural primitives and typical relations between primitives, and should not contain characteristics of the misestimated numeral. Thereby, this type of errors should also be reduced by the verifier. Moreover, errors in Category 3 can be corrected by implementing a classifier based on structural features. Since their misclassified samples have perfect structural primitives and typical shapes, it will not be tough for a classifier to recognize them based on structural features. However, since they were misclassified once, to avoid making another mistake, the recognition threshold should be set at very specific ranges.

Another possible solution of error reduction is combination of multiple classifiers. High reliability is extremely difficult to achieve by a single classifier. So it is reasonable to combine several classifiers to produce a higher accuracy. In recent years, a lot of combination methods have been implemented using different strategies to improve the recognition rate, such as majority voting schemes, statistical approaches, formulations based on Bayesian and Dempster–Shafer theories of evidence. In general, using a combination of classifiers usually yields an improvement in reliability. In our analysis of misclassified data, as shown in Table 3 above, only 2.73% of errors are common to all the classifiers found in this study. These errors can not be eliminated by any approaches of combination of the multitude of classifiers included in this study. However, 60.94% of errors are unique to a classifier and about 80% of errors are produced by 1 or 2 out of 5 classifiers. This indicates that 80% of errors can be reduced by a simple combination method: majority vote is applied. If a thorough analysis of the advantage, limitation and complementary nature of each classifier is carried out and followed by a suitable combination decision rules, a larger error reduction can be accomplished. This observation is one motivation to combine multiple classifiers to achieve a higher accuracy. A hybrid multiple classifier system can be used to achieve this goal. This system effectively integrates cooperation and combination of multiple classifiers. Cooperation is a serial topology of multiple classifiers. For efficiency, Classifiers with high accuracy are used in cooperation with restrict rejection threshold. The patterns rejected by these classifiers are sent to the combination of multiple classifiers, which is a parallel topology. The weights of different classes in each classifier are calculated from the overall performance of each classifier on the training set.

5. Conclusion

In this paper, we have presented a thorough investigation of misclassified handwritten digits by a variety of classifiers produced by numerous research teams located in different parts of the world. The goal of this study is to identify systematically the reasons of misclassification, and to explore possible solutions to reduce errors and improve the reliability of the recognition system. Two types of analysis have been conducted. The first type divided the misclassified data into three categories based on their natures. It also concluded that the majority of errors can be easily recognized by humans, which indicated the possibility of further error reduction. The second type of analysis focused on the common errors made by three or more out of five classifiers and investigated the reasons of misclassification.

Finally, based on the analyses described in this study, two promising ways to reduce errors and improve the system reliability have been proposed: implementation of a verification module and combination of multiple classifiers. These approaches are being studied at CENPARMI.

Acknowledgements

The assistance of J.X. Dong, C.L. He, and L.S. Oliveira in the project is very much appreciated. This project was supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix A

This appendix presents the misclassified data of various classifiers. Column A presents the sequence number of the misclassified sample in the database. Column B presents the misclassification: the first digit is the true identity of the sample; the second digit is the output produced by the classifier.

A	B	A	B	A	B	A	B	A	B
GPR (MNIST database)									
248	4 → 2	1622	0 → 6	3074	1 → 2	3074	1 → 2	7217	0 → 6
321	9 → 8	1879	8 → 3	3521	6 → 4	3521	6 → 4	7260	8 → 2
446	6 → 0	1902	9 → 4	3763	6 → 8	3763	6 → 8	8021	1 → 4
448	4 → 9	2054	4 → 9	3809	7 → 8	3809	7 → 8	8317	7 → 2
675	5 → 3	2071	7 → 9	3870	9 → 4	3870	9 → 4	8326	0 → 8
727	7 → 4	2119	6 → 0	4016	9 → 5	4016	9 → 5	8333	9 → 7
814	9 → 8	2131	4 → 9	4079	9 → 2	4079	9 → 2	8409	8 → 5
939	3 → 5	2136	6 → 1	4177	2 → 7	4177	2 → 7	9016	7 → 2
948	8 → 9	2183	1 → 2	4202	1 → 9	4202	1 → 9	9025	7 → 2
1015	6 → 5	2267	1 → 5	4225	9 → 7	4225	9 → 7	9281	8 → 5
1115	3 → 8	2388	9 → 1	4239	7 → 3	4239	7 → 3	9506	7 → 2

Appendix A (*continued*)

A	B	A	B	A	B	A	B	A	B
1227	7 → 2	2448	4 → 9	4266	4 → 7	4266	4 → 7	9531	9 → 8
1233	9 → 4	2455	6 → 8	4272	5 → 3	4272	5 → 3	9588	9 → 4
1248	9 → 5	2463	2 → 0	4285	9 → 5	4285	9 → 5	9621	9 → 7
1300	5 → 3	2824	7 → 4	4307	3 → 7	4307	3 → 7	9635	0 → 8
1320	8 → 3	2928	3 → 2	4361	5 → 3	4361	5 → 3	9665	2 → 7
1404	1 → 6	2940	9 → 5	4370	9 → 4	4370	9 → 4	9730	5 → 6
1554	9 → 3	2946	3 → 7	4383	4 → 9	4383	4 → 9	9906	3 → 7
1612	3 → 5	3061	9 → 3	4508	1 → 3	4508	1 → 3		
VSVM[b] (MNIST database)									
248	4 → 6	1248	9 → 5	1248	9 → 5	4202	1 → 7	8317	7 → 2
583	8 → 2	1261	7 → 1	1261	7 → 1	4444	3 → 2	8409	8 → 5
584	2 → 7	1879	8 → 3	1879	8 → 3	4498	8 → 7	9506	7 → 2
939	3 → 5	1902	9 → 4	1902	9 → 4	4762	9 → 8	9730	5 → 6
948	8 → 9	2036	5 → 3	2036	5 → 3	4824	9 → 4	9793	4 → 9
1015	6 → 5	2071	7 → 9	2071	7 → 9	5655	7 → 2	9840	2 → 7
1227	7 → 2	2131	4 → 9	2131	4 → 9	5938	5 → 3		
1233	9 → 4	2136	6 → 1	2136	6 → 1	6577	7 → 1		
VSV2 (MNIST database)									
448	4 → 9	1320	8 → 0	2183	1 → 2	3535	4 → 8	6784	1 → 6
583	8 → 2	1531	8 → 7	2294	9 → 6	3559	5 → 0	8326	0 → 6
660	2 → 7	1550	4 → 6	2489	2 → 4	3605	7 → 0	8409	8 → 5
675	5 → 3	1682	3 → 7	2655	6 → 1	3763	6 → 8	9665	2 → 7
727	7 → 3	1710	9 → 5	2928	3 → 2	3870	9 → 4	9730	5 → 6
948	8 → 9	1791	2 → 8	2940	9 → 7	3986	9 → 4	9750	5 → 6
1015	6 → 5	1902	9 → 4	2954	3 → 5	4079	9 → 3	9793	4 → 9
1113	4 → 6	2036	5 → 3	3031	6 → 8	4762	9 → 4	9851	0 → 6
1227	7 → 2	2071	7 → 9	3074	1 → 2	4824	9 → 4		
1233	9 → 4	2099	2 → 0	3226	7 → 9	5938	5 → 3		
1248	9 → 5	2131	4 → 9	3423	6 → 0	6577	7 → 1		
1300	5 → 7	2136	6 → 1	3521	6 → 4	6598	0 → 7		
LeNet5 (MNIST database)									
248	4 → 6	1791	2 → 7	2928	3 → 2	4576	4 → 2	7435	4 → 9
450	3 → 5	1879	8 → 3	2940	9 → 5	4602	8 → 4	8060	2 → 1
583	8 → 2	1902	9 → 4	3031	6 → 0	4741	3 → 5	8095	2 → 8
660	2 → 1	1956	8 → 2	3423	6 → 0	4957	8 → 4	8409	8 → 5
675	5 → 3	2036	5 → 3	3521	6 → 0	5737	6 → 5	8528	4 → 9
830	4 → 8	2044	4 → 8	3763	6 → 8	5750	8 → 5	9010	7 → 2
927	2 → 8	2110	3 → 9	3781	4 → 6	5956	3 → 8	9016	7 → 2
939	3 → 5	2119	6 → 0	3809	7 → 3	5974	3 → 8	9680	6 → 5
1015	6 → 5	2130	9 → 8	3870	9 → 4	6072	9 → 8	9693	9 → 7
1040	7 → 3	2131	4 → 9	3942	4 → 6	6102	1 → 5	9699	6 → 1
1233	9 → 4	2136	6 → 1	4177	2 → 7	6174	9 → 8	9730	5 → 6
1320	8 → 0	2294	9 → 4	4225	9 → 7	6539	6 → 3	9771	5 → 0
1329	7 → 8	2388	9 → 1	4266	4 → 3	6558	0 → 2	9793	4 → 9
1394	5 → 3	2415	9 → 4	4370	9 → 4	6559	6 → 5	9905	2 → 8
1531	8 → 7	2463	2 → 0	4406	9 → 4	6593	9 → 5		
1622	0 → 6	2655	6 → 1	4426	9 → 4	6598	0 → 7		
1682	3 → 7	2922	3 → 5	4498	8 → 7	6784	1 → 6		

(*continued on next page*)

Appendix A (*continued*)

A	B	A	B	A	B	A	B	A	B
VSVM (CENPARMI database)									
397	$1 \to 2$	809	$4 \to 6$	1453	$7 \to 9$	1663	$8 \to 9$	1965	$9 \to 8$
588	$2 \to 8$	812	$4 \to 6$	1520	$7 \to 4$	1668	$8 \to 0$	1994	$9 \to 2$
648	$3 \to 9$	919	$4 \to 9$	1525	$7 \to 4$	1764	$8 \to 9$		
747	$3 \to 2$	1140	$5 \to 8$	1527	$7 \to 2$	1882	$9 \to 7$		
783	$3 \to 2$	1351	$6 \to 0$	1637	$8 \to 5$	1928	$9 \to 8$		
790	$3 \to 2$	1452	$7 \to 9$	1640	$8 \to 2$	1952	$9 \to 8$		
VSVM (USPS database)									
18	$6 \to 4$	18	$6 \to 4$	18	$6 \to 4$	18	$6 \to 4$	1814	$1 \to 4$
28	$3 \to 5$	28	$3 \to 5$	28	$3 \to 5$	28	$3 \to 5$	1815	$1 \to 7$
53	$1 \to 5$	53	$1 \to 5$	53	$1 \to 5$	53	$1 \to 5$	1816	$1 \to 4$
79	$2 \to 5$	79	$2 \to 5$	79	$2 \to 5$	79	$2 \to 5$	1865	$9 \to 8$
165	$0 \to 8$	165	$0 \to 8$	165	$0 \to 8$	165	$0 \to 8$	1872	$4 \to 7$
199	$8 \to 0$	199	$8 \to 0$	199	$8 \to 0$	199	$8 \to 0$	1952	$5 \to 8$
234	$1 \to 6$	234	$1 \to 6$	234	$1 \to 6$	234	$1 \to 6$	1978	$5 \to 3$
266	$4 \to 7$	266	$4 \to 7$	266	$4 \to 7$	266	$4 \to 7$		
340	$7 \to 4$	340	$7 \to 4$	340	$7 \to 4$	340	$7 \to 4$		
485	$3 \to 5$	485	$3 \to 5$	485	$3 \to 5$	485	$3 \to 5$		
POE (MNIST database)									
152	$9 \to 8$	1791	$2 \to 7$	3031	$6 \to 0$	4285	$9 \to 5$	6174	$9 \to 8$
248	$4 \to 6$	1801	$6 \to 4$	3061	$9 \to 7$	4345	$9 \to 5$	6533	$0 \to 5$
265	$9 \to 4$	1869	$1 \to 2$	3074	$1 \to 2$	4438	$3 \to 2$	6556	$8 \to 9$
322	$2 \to 7$	1872	$2 \to 8$	3290	$8 \to 9$	4478	$0 \to 6$	6569	$9 \to 7$
360	$9 \to 4$	1879	$8 \to 3$	3331	$2 \to 8$	4549	$5 \to 6$	6577	$7 \to 1$
543	$8 \to 5$	1902	$9 \to 4$	3337	$5 \to 9$	4640	$8 \to 9$	6598	$0 \to 9$
583	$8 \to 3$	1904	$7 \to 2$	3476	$3 \to 7$	4658	$3 \to 2$	6652	$0 \to 8$
584	$2 \to 8$	2044	$4 \to 8$	3504	$9 \to 3$	4672	$8 \to 3$	6756	$8 \to 9$
660	$2 \to 7$	2099	$2 \to 0$	3512	$2 \to 7$	4700	$6 \to 1$	7122	$8 \to 9$
685	$7 \to 3$	2110	$3 \to 7$	3521	$6 \to 4$	4741	$3 \to 5$	7801	$3 \to 2$
741	$4 \to 9$	2119	$6 \to 0$	3535	$4 \to 8$	4762	$9 \to 7$	7822	$3 \to 2$
727	$7 \to 3$	2130	$9 \to 2$	3559	$5 \to 0$	4764	$5 \to 6$	7922	$8 \to 0$
948	$8 \to 9$	2136	$6 \to 1$	3598	$9 \to 3$	4808	$8 \to 0$	8060	$2 \to 1$
1015	$6 \to 5$	2183	$1 \to 3$	3605	$7 \to 2$	4824	$9 \to 4$	8070	$2 \to 1$
1040	$7 \to 3$	2186	$0 \to 8$	3703	$5 \to 3$	4887	$7 \to 1$	8095	$2 \to 8$
1113	$4 \to 6$	2294	$9 \to 4$	3719	$4 \to 9$	5047	$3 \to 5$	8096	$4 \to 1$
1182	$6 \to 1$	2407	$9 \to 4$	3752	$7 \to 2$	5279	$8 \to 7$	8113	$2 \to 8$
1183	$6 \to 5$	2455	$6 \to 8$	3767	$3 \to 5$	5289	$8 \to 7$	8247	$3 \to 9$
1227	$7 \to 2$	2463	$2 \to 0$	3768	$7 \to 3$	5332	$1 \to 6$	8278	$3 \to 8$
1233	$9 \to 4$	2489	$2 \to 4$	3797	$2 \to 8$	5458	$1 \to 8$	8297	$3 \to 8$
1248	$9 \to 5$	2575	$5 \to 3$	3809	$7 \to 8$	5635	$2 \to 3$	8326	$0 \to 6$
1261	$7 \to 1$	2583	$9 \to 7$	3860	$9 \to 4$	5643	$1 \to 8$	8398	$3 \to 5$
1291	$3 \to 5$	2598	$5 \to 3$	3894	$5 \to 6$	5677	$4 \to 2$	8409	$8 \to 5$
1320	$8 \to 3$	2619	$3 \to 5$	3903	$5 \to 3$	5737	$6 \to 0$	8521	$4 \to 9$
1365	$8 \to 2$	2655	$6 \to 1$	3907	$1 \to 3$	5888	$7 \to 0$	8528	$4 \to 9$
1394	$5 \to 3$	2761	$9 \to 4$	3942	$4 \to 6$	5938	$5 \to 3$	9016	$7 \to 2$
1396	$2 \to 8$	2772	$4 \to 9$	3969	$5 \to 3$	5956	$3 \to 8$	9025	$7 \to 2$
1523	$7 \to 9$	2897	$8 \to 0$	3996	$3 \to 5$	5973	$5 \to 3$	9665	$2 \to 7$
1528	$1 \to 5$	2922	$3 \to 8$	4066	$0 \to 2$	5982	$5 \to 3$	9730	$5 \to 6$
1531	$8 \to 7$	2940	$9 \to 5$	4079	$9 \to 3$	5998	$5 \to 3$	9746	$4 \to 7$
1550	$4 \to 6$	2953	$3 \to 5$	4177	$2 \to 7$	6082	$9 \to 3$	9840	$2 \to 0$
1560	$9 \to 3$	2954	$3 \to 5$	4200	$7 \to 9$	6092	$9 \to 3$	9851	$0 \to 6$
1682	$3 \to 7$	2971	$5 \to 3$	4202	$1 \to 4$	6167	$9 \to 3$		
1710	$9 \to 3$	3006	$9 \to 4$	4225	$9 \to 7$	6169	$9 \to 3$		

Appendix A (*continued*)

A	B	A	B	A	B	A	B	A	B
MLP (NIST SD19)									
4177	0 → 4	593	1 → 2	5869	3 → 5	5869	3 → 5	420	8 → 0
4179	0 → 4	594	1 → 2	168	3 → 7	168	3 → 7	421	8 → 0
4358	0 → 4	5025	1 → 7	32	3 → 7	32	3 → 7	5825	8 → 0
5447	0 → 4	5026	1 → 7	3965	3 → 7	3965	3 → 7	1234	8 → 1
5334	0 → 6	514	1 → 7	3972	3 → 7	3972	3 → 7	2801	8 → 2
4945	0 → 8	516	1 → 7	3975	3 → 7	3975	3 → 7	5867	8 → 2
4180	0 → 9	5765	1 → 7	2890	3 → 8	2890	3 → 8	1759	8 → 3
3375	1 → 2	6103	1 → 7	3599	3 → 8	3599	3 → 8	1850	8 → 4
359	1 → 2	6105	1 → 7	5196	3 → 9	5196	3 → 9	2396	8 → 4
360	1 → 2	2813	1 → 8	1833	4 → 6	1833	4 → 6	1238	8 → 9
361	1 → 2	1613	2 → 3	3398	4 → 7	3398	4 → 7	3915	9 → 3
4560	1 → 2	444	2 → 7	968	4 → 9	968	4 → 9	1296	9 → 4
4564	1 → 2	1290	2 → 8	160	5 → 1	160	5 → 1	1300	9 → 4
4569	1 → 2	2666	2 → 8	2104	5 → 3	2104	5 → 3	1302	9 → 4
4959	1 → 2	2919	2 → 8	494	5 → 3	494	5 → 3	1685	9 → 4
5136	1 → 2	5218	2 → 8	4206	5 → 6	4206	5 → 6	2544	9 → 4
5137	1 → 2	2819	3 → 0	413	5 → 8	413	5 → 8	2730	9 → 4
5141	1 → 2	840	3 → 0	157	5 → 9	157	5 → 9	2805	9 → 4
5145	1 → 2	4570	3 → 1	3229	5 → 9	3229	5 → 9	3165	9 → 4
5148	1 → 2	3600	3 → 2	813	5 → 9	813	5 → 9	4663	9 → 4
5601	1 → 2	4417	3 → 2	815	5 → 9	815	5 → 9	4785	9 → 4
586	1 → 2	5692	3 → 2	516	5 → 9	516	5 → 9	3052	9 → 5
588	1 → 2	1567	3 → 5	517	5 → 9	517	5 → 9	270	9 → 7
590	1 → 2	3411	3 → 5	818	5 → 9	818	5 → 9		

References

Ahmed, P., Suen, C.Y., 1984. Edge classification and extraction of shape features. In: Proceedings of ICPR, Montreal, Canada, pp. 593–596.

Bromley, J., Sackinger, E., 1991. Neural-network and *k*-nearest-neighbor-classifiers. Technical Report 11359-910819-16TM, AT&T.

Decoste, D., Scholkopf, B., 2002. Training invariant support vector machines. Machine Learning 46, 161–190.

Dong, J.X., 2003. Speed and accuracy: large-scale machine learning algorithms and their applications. Ph.D. Thesis, Department of Computer Science, Concordia University.

Dong, J.X., Krzyzak, A., Suen, C.Y., 2001. Statistical result of human performance on USPS database. Technical Report, CENPARMI, Concordia University.

Dong, J.X., Krzyzak, A., Suen, C.Y., 2002. Local learning framework for handwritten character recognition. Eng. Appl. Artificial Intell. 15, 151–159.

Kussul, E., Baidyk, T., 2003. Permutative coding technique for handwritten digit recognition system. In: Proc. International Joint Conference on Neural Networks, vol. 3, Portland, Oregon, USA. pp. 2163–2168.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 80 (11), 2278–2324.

Liu, C.L., Nakashima, K., Sako, H., Fujisawa, H., 2004. Handwritten digit recognition: Investigation of normalization and feature extraction techniques. Pattern Recognition 37, 265–279.

Mayraz, G., Hinton, G.E., 2002. Recognizing handwritten digits using hierarchical products of experts. IEEE Trans. Pattern Anal. Machine Intell. 24 (2), 189–197.

Oliveira, L.S., Sabourin, R., Bortolozzi, F., Suen, C.Y., 2003. Impacts of verification on a numeral string recognition system. Pattern Recognition Lett. 24, 1023–1031.

Takahashi, H., Griffin, T., 1993. Recognition enhancement by linear tournament verification, In: Proc. 2nd International Conference on Document Analysis and Pattern Recognition, vol. 1, Tsukuba, Japan, pp. 585–588.

Tan, J.N., 2004. Automatic verification of the outputs of multiple classifiers for unconstrained handwritten numerals. Master thesis, Department of Computer Science, Concordia University.