MULTIPLE CLASSIFIERS ERROR RATE OPTIMIZATION APPROACHES OF AN AUTOMATIC SIGNATURE VERIFICATION (ASV) SYSTEM

Sharifah M. Syed Ahmad
Universiti Tenaga Tenaga Nasional (UNITEN), 43009 Kajang, Selangor, Malaysia

Keywords: Multiple classifiers, Borda Count, Equal Error Rate (EER), Automatic Signature Verification.

Abstract: Decision level management is a crucial aspect in an Automatic Signature Verification (ASV) system, due to its nature as the centre of decision making that decides on the validity or otherwise of an input signature sample. Here, investigations are carried out in order to improve the performance of an ASV system by applying multiple classifier approaches, where features of the system are grouped into two different sub-sets, namely static and dynamic sub-sets, hence having two different classifiers. In this work, three decision fusion methods, namely Majority Voting, Borda Count and cascaded multi-stage cascaded classifiers are analyzed for their effectiveness in improving the error rate performance of the ASV system. The performance analysis is based upon a database that reflects an actual user population in a real application environment, where as the system performance improvement is calculated with respect to the initial system Equal Error Rate (EER) where multiple classifiers approaches were not adopted.

1 INTRODUCTION

There is a wide diversity of information available to characterize human signatures, which offers an opportunity to create an Automatic Signature Verification (ASV) system with a high degree of flexibility and sophistication. For example an ASV system can be built out of static and dynamic features (Lee et al., 1996), heuristics and spectral based features (Allgrove and Fairhursty), global and local features (Sansone and Vento, 2000), as well as parametric and non-parametric features (Kegelmeyer and Bower, 1997). However, it is probably ineffective in terms of the error rate to combine disparate forms of signature information in a single classifier, due to the possibility of features incompatibility that is caused by different nature of features. Thus, here, multiple classifiers approaches are adopted, where each individual classifier operates on of a pool of features of the same type, and the overall system is built out of a combination of these classifiers.

The rationale for adopting a multiple classifier approach in increasing the performance of an ASV system is that such an approach makes it possible to compensate for the weakness of each individual classifier while preserving its own strengths (Sansone and Vento, 2000), (Allgrove and Fairhursty). Kittler (Kittler et al., 1998), (Kittler, 1999) in his research has pointed out that in order to achieve system performance improvement, the system should be built up from a set of highly complementary classifiers in terms of error distribution. This means that the classifiers should not be strongly related in their miss-clarification, indicating a requirement for selecting classifiers that are error- independent of each other.

However, some studies have questioned the need of such classifier error independence in obtaining system improvement (Demirecler and Altincay, 2002), (Kuncheva et al., 2000). Demirekler and Altincay (Demirecler and Altincay, 2002) in their investigations have noted that independent multiple classifier may not necessarily yields the best system performance. Therefore, the investigation on error distribution between different classifiers prior to the implementation of the multiple classifiers approach is not considered here. The only prior work carried out is the analysis of individual classifier performance, which is discussed in section 4.

M. Syed Ahmad S. (2007).
MULTIPLE CLASSIFIERS ERROR RATE OPTIMIZATION APPROACHES OF AN AUTOMATIC SIGNATURE VERIFICATION (ASV) SYSTEM.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 257-263

2 USER DATABASE

The work reported here is based on a test set that is aligned with the guidance of good practice drafted by the UK Biometric Working Group which requires scenario evaluation to be carried out on a database that reflects an actual user population in a real application environment (August, 2002). The database used in the verification was compiled as part of the KAPPA Project in January, 1994 where a large cross section of the general public were invited to take part in data collection trials carried out at a major Post Office branch over a few months period of time (Canterbury, 1994).

More than 5000 samples were collected, where the minimum number of samples submitted per subject is 3, and some users are reported to have submitted more than 40 samples. Each user is assigned with a unique identifier (ID) number, where all samples of the same individual are stored under the same user ID entry. The data collected do not include any attempted forgeries, which do not allow for investigations on forged signatures. Signature samples were captured using a digitizing tablet, connected to a PC. The tablet captured information in the form of signing co-ordinates and timing data, as well as pen-up or pen-down data indication.

3 SYSTEM DESIGN AND TEST METHODOLOGY

A relatively simple straightforward prototype system is constructed based on a pool of 22 global normalized features, consisting of 8 static and 14 dynamic features (as listed in Table 1 and Table 2 respectively), a simplification of Cornell ASV system (Lee et al., 1996).

Table 1: The Static Classifier's Feature Set.

Feature	Feature Description
S1	Normalized [first x - max. x]
S2	Normalized [first x - min. x]
S3	Normalized [last x - max. x]
S4	Normalized [last x - min. x]
S5	Normalized [first y - max. y]
S6	Normalized [first y - min. y]
S7	Normalized [last y -max. y]
S8	Normalised [last y - min. y]

Though the database adopted is built in an online mode that captured dynamic information of the signing operation, the testing is carried out purely using offline batch processing. Hence the test results are repeatable for the same test scenario due to the fixed nature of the database.

Table 2: The Dynamic Classifier's Feature Set.

Feature	Feature Description
D1	Normalized time
D2	Normalized max. speed
D3	Avg. speed / max. speed
D4	Normalized x zero velocity
D5	Normalized x positive velocity
D6	Normalized x negative velocity
D7	Normalized y zero velocity
D8	Normalized y positive velocity
D9	Normalized y negative velocity
D10	Avg. speed / max x velocity
D11	Avg. speed / max y velocity
D12	Min. x velocity / avg. x velocity
D13	Min. y velocity / avg. y velocity
D14	Normalized min. acceleration

The evaluation is carried out on two different modes, as defined by the UK Biometric Working Group (August, 2002), which are:

- **genuine claim of identity**
  A test is carried out to compare a user signature sample with his / her genuine reference data under the same ID. Hence any invalid verification gives rise to False Rejection Rate (FRR).
- **impostor claim of identity**
  A test is carried out to compare a user signature sample with a different subject reference data under different IDs. Hence any valid verification gives rise to False Acceptance Rate (FAR). Formally, such 'forgeries' are unskilled, with no deliberate attempt to reproduce another person's signature. This type of forgery is known as 'random' forgery (Gnee, 2000). Another type of forgery which is 'skilled' forgery is not considered here due to the unavailability of forged signature samples.

At the initial stage, a system prototype (i.e. without the implementation of multiple classifiers that combines all features in the ASV system) is created in order to provide for testing as well as benchmarking for the envisaged optimization investigations. In the verification process of the system prototype, each feature cast a binary accept or reject vote, whereas the validity of a genuine signature sample is decided based on the number of accumulated accept votes cast by all features that is compared against an overall *system_threshold* (i.e. the ASV system operate based on a threshold voting mechanism). This *system_threshold* is adjusted until the ASV system reached its Equal Error Rate (EER), which is recorded here occurred at 8.31%. Thus, the 8.31% EER figure remains the analysis benchmark when comparing the system performance with the implementation of multiple classifiers approaches.

4 INDIVIDUAL CLASSIFIER PERFORMANCE ANALYSES

For the individual classifier performance analyses, the features are grouped based on the static / dynamic nature of each feature. For each classifier *i*, where *i* is the classifier index, a corresponding classifier threshold (i.e. *classifier_i_threshold*) is maintained which controls the degree of classifier *i*'s sample acceptance. Here, the total number of accepted features in classifier *i* is compared against the *classifier_i_threshold*, which follows the classification rule in equation 1. The *classifier_i_threshold* is adjusted until the classifier Equal Error Rate (EER) is achieved.

If (*x_total_classifier_i_feature accepted* >= *classifier_i_threshold*)
Then tested signature is assumed to be genuine; otherwise it is assumed as a forgery (1)

Table 3: Results of Individual Classifiers' Performance Analyses.

Estimated EER (%) for Dynamic Classifier	**Estimated EER (%) for Static Classifier**
9.93	16.48

The result of individual classifier performance analyses is as listed in Table 3. The figures indicate that the dynamic classifier in this ASV system that is built up from a sub-set of 14 dynamic features has higher identifying and / or discriminating capability compared to that of the static classifier, which is built up from a sub-set of 8 features. This information regarding the non-uniformity of the classifiers' performance provides some insights on the selection of a suitable combining strategy in the subsequent envisaged analyses of section 5. The results in table 3 also show that both classifiers perform poorer on their own, as compared to the performance of the prototype system which combines all features which is recorded at 8.31% of the system EER. Thus, the next step is to investigate the possibility of increasing the ASV system error rate performance when the decisions cast by both static and dynamic classifiers are combined based on several multiple classifier approaches.

5 SYSTEM OPTIMIZATION BY USING MULTIPLE CLASSIFIERS APPROACHES

The main concern in using multiple classifiers is selecting the combining strategy, a process what is commonly referred to as 'decision fusion'. Decision fusion can be defined as a formal framework that expresses the means and tools for the integration of data originating from different classifiers (Arif and Vincent, 2003). For the last few decades, there has been extensive research in the various types of decision fusion methods. Clearly there is no one universally best combiner existing that would suit all applications. A study carried out by Demirekler and Altincay (Demirecler and Altincay, 2002) has discussed the impact of decision fusion method on the choice of classifier. For example, for a given set of classifiers *N*, there is a sub-set of class1ifiers *N'* that yields the best performance under a given combination rule *c*, where an addition of another classifier will only degrade the system performance. Lei, Krzyzak and Suen (Lei et al., 1992) in their research have defined three types of multiple classifiers combining strategy based on the classifiers' output:

- combination is made according to the output of a unique label
  Each classifier produces an output label (e.g. a signature sample is accepted or rejected by classifier *i*, where *i* is the classifier index), which is processed and integrated by the system in order to produce a final label
- combination is made based on ranking information
  Each classifier produces a ranked set of labels (e.g. a signature sample that has high probability of being genuine is then given a high acceptance rank, where else a signature sample that has a high probability of being forged is then given a low acceptance rank), and the ranking of the labels is added up and processed at the final stage.
- combination is made according to measurements output
  Each classifier produces measurement level of information, instead of definite decision (e.g. the acceptance degree of an input signature sample), which is processed and integrated by the system in order to produce the final decision.

In this work, several decision fusion methods that represent each of the above said decision combiner types are analyzed for their effectiveness in improving the error rate performance of the ASV system. These include majority voting, Borda Count and multi-stage cascaded classifiers respectively.

5.1 System Optimization by Using Majority Voting

In a voting algorithm, each classifier is required to produce a soft decision (i.e. a vote) which is added up and analyzed in order to arrive at a hard decision about an input sample (Lam and Suen, 1997). In an Automatic Signature Verification (ASV) system, the decision takes the form of either one of the two verification classes, namely 'sample accepted' or 'sample rejected'.

One of the main advantages of a voting system is that it allows for a wide variety of classifier types to be combined without much concern about the underlying classifier processing methodologies. Basically, it treats individual classifiers as 'black boxes' and needs no internal information on the implementation (Lin et al., 2003), which in turn offers a great deal of system flexibility, generality and simplicity.

There are many different forms of voting system such as threshold voting, majority voting, plurality voting and others. An example of a threshold voting system is the decision-making mechanism applied in this ASV prototype system where a signature sample is classified as genuine only if the total number of acceptance votes exceeds the defined system threshold. On the other hand, both majority voting and plurality voting systems require more than half of the candidates' votes and most votes respectively (Lin et al., 2003). However, in a two-class recognition problem, a plurality voting mechanism operates in a similar way to that of a majority voting mechanism (i.e. the class with the most votes also receives more than half of the candidates' votes). Therefore, the literature on pattern recognition generally does not differentiate between these two voting systems. Here, such a mechanism is referred to as a majority voting approach in order to avoid confusion.

Since the ASV system is built up of only two classifiers (i.e. static and dynamic classifier), in order to satisfy the requirement of the majority voting definition, the final decision is based on whichever class that receives two votes (i.e. more than 1 which is more than half the number of the total votes). However, a conflict exists when both classes obtained equal number of votes (i.e. one vote each). Demirekler and Altincay (Demirecler and Altincay, 2002) suggested that in order to resolve the conflict, a random decision is selected among these classes. However such an approach is not considered here due to the uncertainty of random decision accuracy. Instead a probably more realistic approach is to resolve the conflict based on the statistical individual classifier performance analysis that is carried out previously in section 4. Here, since the dynamic classifier has a lower Equal Error Rate (EER), the resolving hard decision is based on the soft decision of the dynamic classifier. The rationale behind this is that the dynamic classifier has lower probability of making a classification error, and thus is more accurate and more reliable compared to the static classifier.

5.2 System Optimization by Using Borda Count

In a voting system, each classifier is required to cast a specific vote amongst one of the available accept-reject classes. This can be 'inaccurate' in the case of sample status is uncertain, and such errors will have major impacts especially in a two-class verification situation, where the probability of error is high, that is 0.5. A Borda Count approach is capable of overcoming the 'lack of depth' problem of the voting system, by allowing each classifier to rank the classes indicating which the more likely candidates are. This approach was presented in 1770 by Jean-Charles of Borda, hence the name 'Borda Count' (Arif and Vincent, 2003). According to him, the class with highest recognition probability receives the highest rank. Consequently, at the decision level, the ranks are added up and the final decision is decided based on the class with the highest accumulated rank.

Erp and Schomaker (Erp and Schomaker, 2000) have analysed the effectiveness of Borda count in a great detail, and they claimed that it is an easy and a powerful method in combining rankings. They have also highlighted the possible limitation of this approach which results may be susceptible to extreme voting by some classifiers. Nevertheless, here, the Borda Count approach is adjusted, where the classes are divided into three categories which are:

- class A - class 'sample accepted'
- class U - class 'sample status is uncertain'
- class R - class 'sample rejected'

Here, since there are three classes (i.e. m is 3), therefore the class with the highest recognition probability receives three ranks (i.e. highest ranks equal to *m* that is 3). The following alternative receives one less rank than the other (i.e. *m*-1=2, *m*-2=1). Therefore all classes will at least receive 1 rank from each classifier. The ranking is based on the number of accepted features within a specified range. The threshold for each range is adjusted until the system EER is achieved. Figure 1 illustrates the ranking mechanism of the designed Borda Count approach.

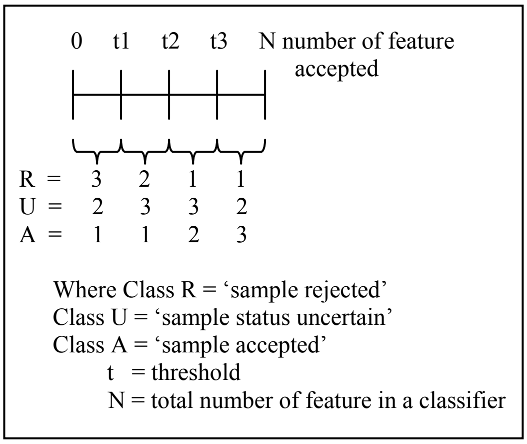

Figure 1: Borda Count Ranking Method.

However, there is also a need in the Borda count approach for a mechanism to solve the conflict of having two classes with an equal number of highest accumulated ranks, similar to the case that needed to be addressed by implementing the majority voting approach. Here, when class U is tied with class A or class R, then the decision will be based on the class that is tied to class U. In addition to this, the system should also deal with the case when class U receives the highest accumulated ranks, since the final system decision can either be in the form of sample accepted or rejected. The solution applied here is simple, that is to be based on the class that receives the second highest mode. Should there be a tie between class R and class A, then the decision will be based on the results of the dynamic classifier, due to the lower EER of the dynamic classifier as compared to that of the static classifier (i.e. as analyzed in section 4). Here, an acceptance classifier threshold is maintained for the dynamic classifier that decides whether the sample is accepted or rejected.

5.3 System Optimization by Using Multi-Stage Cascaded Classifiers

In section 5.1 and section 5.2, the multiple classifier ASV system using the Majority Voting algorithm and the Borda count method as decision fusion strategies have been designed in order to analyze their effectiveness in optimizing the system error rate performance. For both methods, the vote / rank cast by each classifier is treated equally in a parallel decision making process, except for the mechanisms designed to resolve conflicts in handling ties in the case of equal number of highest votes / accumulated ranks between sample class rejected and sample class accepted. However, such approaches lack an ability to differentiate between different individual classifier's performances, where clearly, as indicated in section 4, the dynamic classifier has lower recognition error rate compared to that of the static classifier. Here, an analysis of a multi-stage cascaded classifiers system is performed, where the soft decisions of individual classifiers are taken into account one after another in a pipeline prioritized nature. A theoretical discussion of such an approach is discussed in a detail by Pudil, Novovicova, Blaha, and Killter (Pudil et al., 1992).

Sansone and Ventro (Sansone and Vento, 2000) in their studies have investigated the effectiveness of a three-stages cascaded multiple classifier system in an Automatic Signature Verification (ASV) system. They have grouped the features into two classifiers where each classifier is designed to tackle one type of forgery at a time. The first stage consists of a simple classifier that is devoted to eliminate random forgeries, where only signature samples that pass the first stage will be forwarded to the second stage that handles skilled forgeries. If the system fails to detect forgeries, then the sample is forwarded to the third stage that takes into account the confidence level of the previous two stages' decisions. The rationale behind such an approach is that most classifiers are only good at detecting one type of forgery, and that the performance of the system decreases when attempting to eliminate all types of forgery simultaneously. Hence, the classifiers are arranged in a cascaded nature in order to handle one type of forgery at a time.

Though this allows for higher system efficiency in handling different types of forgery, it does not necessarily increase the performance of the system in terms of error rate since the whole system decision making can be summarized as a parallel logical 'AND' function (i.e. a signature sample is accepted if and only if the signature sample pass all three stages). Furthermore, since the test cases for this research do not include test for skilled forgery, such an approach is not suitable to be implemented in this analysis.

Instead, here, the cascaded classifiers are arranged based on individual classifier's performance of section 4. The classifier with the higher performance (i.e. the dynamic classifier) is given a higher priority where its soft decision is considered at the first stage.

In order to not resemble the parallel logical 'AND' function, three classes are maintained for the first stage, which are:

- class A - class 'sample accepted'
- class U - class 'sample status is uncertain'
- class R - class 'sample rejected'

Thus, two feature thresholds are maintained for the dynamic classifier. The first threshold (i.e. *accept_threshold*) is the minimum limit of number of features that must be accepted in order to cast the final decision of sample accepted. The second threshold (i.e. *reject_threshold*) is the maximum limit of the number of features that must be accepted to invoke the final decision of sample rejected. Only the signature samples that fall under class U (i.e. the total number of features accepted lies in between *accept_threshold* and *reject_threshold*) are forwarded to the second stage, otherwise the system will output a final 'sample accepted' or 'sample rejected' decision. The second stage consists of the static classifier. Here, samples whose statuses are undecided in the first stage are processed based on the soft decision of the second stage. The output of the second stage is in the form of a final decision 'sample accepted' or 'sample rejected'.

6 EXPERIMENT RESULTS

The results on all three multiple classifier approaches' analyses are as shown in Table 4.

Table 4: Results using Multiple Classifier Approaches.

Multiple Classifier Combining Strategy	System EER	% of EER Improvement
(1) Majority voting	9.93	-19.49
(2) Borda Count	8.20	1.32
(3) Multi-stage cascaded classifiers	7.57	8.90

The system improvement is calculated with respect to the original system prototype performance without the use of multiple classifiers, which is at 8.31% of the system EER. The results in Table 4 show that the majority voting approach has negative impacts on the system performance (i.e. the system performance deteriorates). This indicates the failure of the majority voting system as a performance optimization tool for this ASV system. A possible explanation for it is the lack of discrimination within each single classifier (i.e. each classifier has a high level of classifier error rates) that leads to a generalized system response which is highly inaccurate. This may also be caused by the limitation on the number of classifiers involved in the voting system.

Where as, for the Borda Count multiple classifier approach, the system has shows improvement in the overall system error rate, however the figure is relatively small that is around 1.32%. The only multiple classifier approach that have successfully increased the system EER to an acceptable level is the multi-stage cascaded classifier approach. The main advantage of this method as compared to the former two tools is that it recognizes the different error rate performances of different classifiers by evaluating the decision cast by each classifier in a cascaded prioritized manner and not in a parallel equal nature (i.e. one that produced the least error rate is given the highest priority). This could probably suggest its superiority in producing the lowest system error rate amongst the analyzed approaches.

7 CONCLUSIONS

In this study, three different types of multiple classifier approach are analyzed to determine their effectiveness as an alternative to the single classifier system that was previously implemented in the ASV prototype. In a multiple classifier system, several classifiers are maintained, where the soft decisions of each classifier are combined according to a combining criterion. First, in order to execute this, the original pool of 22 features is divided into two sub-sets based on the nature of each feature, namely dynamic and static sub-sets. These in turn are treated as individual classifiers. A classifier performance analysis is carried out prior to the investigation of the multiple classifier approach. Here, a threshold voting algorithm is applied at the classifier level, where the corresponding threshold is adjusted until an Equal Error Rate (EER) is achieved. The results show that the dynamic classifier has lower EER compared to that of the static classifier, hence suggesting that the dynamic information used in this

ASV system is more accurate compared to the static information.

The second step carried out here is the analyses of the effectiveness of three different types of combining strategy which are based on majority voting, Borda Count and a multi-stage cascaded classifier configuration In general the results demonstrate that a multiple cla ssifier approach is a possible optimisation tool for an ASV system. However, not all combining strategies are effective in order to achieve a performance increment. For a system with high individual classifiers error rates, a voting mechanism is unsuitable, due to the inability of individual classifier in determining the exact status of an input sample. Thus, for such a situation, a combining algorithm that allows a classifier to output an 'uncertain' status of a sample is highly desirable. It is also best to choose a combining strategy that acknowledges and treats decisions cast by different classifiers in a prioritized cascaded manner for a situation where different classifiers recorded considerably different error rate performances.

REFERENCES

Communications-Electronics Security Group (CESG), "Best Practices in Testing and Reporting Performance of Biometric Devices", Version 2.01, August 2002, ISSN 1471-0005.

http://www.cesg.gov.uk/technology/biometrics/media/Best%20Practice.pdf

Luan L. Lee, Toby Berger, Erez Aviczer, "Reliable On Line Human Signature Verification Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence (18,6), June 1996.

University of Kent at Canterbury, BT Technology Group Ltd, "KAPPA Summary Report", May 1994.

Sue Gnee Ng, University of Kent at Canterbury, PhD Thesis, "Optimisation Tools for Enhancing Automatic Signature Verification", 2000.

C. Sansone, M. Vento, "Signature Verification: Increasing Performance by a Multi Stage System", Pattern Analysis and Applications, (3) 2000, page 169 – 181.

Mubeccel Demirecler, Hakan Altincay, "Plurality Voting Based Multiple Classifier Systems: Statistically Independent With Respect To Dependent Classifier Sets", Patter Recognition, (35) 2002, page 2365 – 2379.

C. Allgrove, M.C. Fairhurst, University of Kent at Canterbury, Technical Report, "Majority Voting – A Hybrid Classifier Configuration for Signature Verification".

M. Arif, N. Vincent, Francois Rabelais University, Technical Report, "Comparison of 3 Data Fusion Methods for an Offline Signature Verification Problem, 2003.

J. Kittler, Mohamad Hatef, Robert P.W. Duin, Jiri Matas, "On Combining Classifiers", IEEE Transactions on Pattern Analysis and Machine Intelligence, (20, 3) March 1998, page 226 – 239.

J. Kittler, "Combining Classifiers: A Theoretical Framework", Pattern Analysis and Applications, (1) 1999, page 18 – 27.

L.I. Kuncheva, C.J. Whitaker, C.A. Shipp, R.P.W. Duin, "Is Independence Good for Combining Classifiers?", Proceedings of the 15th International Conference on Pattern Recognition, 2000.

X. Lei, A. Krzyzak, H.M. Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition", IEEE Transactions on Systems, Man and Cybernatics, (22, 5) 1992, page 418 – 435.

L. Lam, C.Y. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of Its Behaviour and Performance", IEEE Transaction on Systems, Man and Cybernatics, (27,5) 1997, page 553 – 568.

X. Lin, S. Yacoubi, J. Burns, S. Simske, HP Laboratories, "Performance Analysis of Pattern Classifier Combination by Plurality Voting", 2003. http://citeseer.nj.nec.com/cs

M. V. Erp, L. Schomaker, NICI, Netherlands, "Variants of the Borda Count Method for Combining Ranked Classifier Hypotheses", 2000.

http://citeseer.nj.nec.com/cs

P. Pudil, J. Novovicova, S. Blaha, J. Kittler, "Multistage Pattern Recognition with Reject Option", Proceedings on the 11th International Conference on Pattern Recognition (ICPR), 1992, page 92 – 95.

W. P. Kegelmeyer Jr, K. Bower, "Combination of Multiple Classifiers Using Local Accuracy Estimates", IEEE Transactions on Pattern Analysis and Machine Intelligence (19), April 1997.

Y.S. Huang, C.Y. Suen, "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals", IEEE Transactions on Pattern Analysis and Machine Intelligence, (17,1), January 1995.