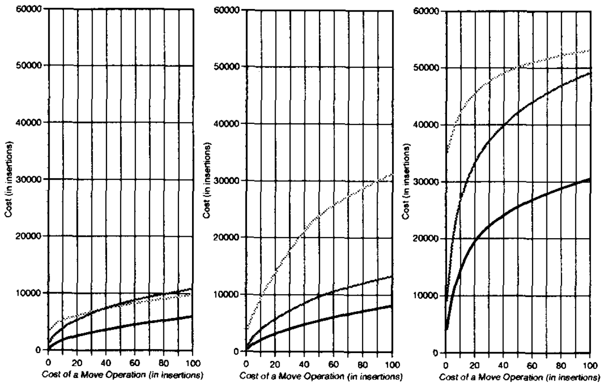
study suggests that the recognition of tables is an important research topic.

Finally, we note that the editing model could be extended to incorporate other edit operations, such as block deletion. In the experiment, we assigned the pages to one of three classes and studied the behavior of the automatic zoning algorithms. We plan to analyze further what kind of layout features make automatic zoning difficult and to study the skew sensitivity of these algorithms.



(a) (b) (c)

Fig. 7. Cost of Correcting Automatic Zoning Errors for the best, middle, and worst OCR system:(a)single column pages, (b) multi-column pages, (c) table pages.

ACKNOWLEDGMENT

This research was funded by the U.S. Department of Energy. We received valuable assistance from Professor S. Latifi (UNLV).

REFERENCES

[1] S. Randriamasy, L. Vincent, and B. Wittner, "An Automatic Benchmarking Scheme for Page Segmentation," *Document Recognition,* SPIE Proceedings Series, vol. 2181, pp. 217-230, 1994.

[2] S. N. Srihari, *Computer Text Recognition and Error Correction,* IEEE Computer Society, 1985.

[3] S. V. Rice, J. Kanai, and T. A. Nartker, "A Difference Algorithm for OCR-Generated Text," *Advances in Structural and Syntactic Pattern Recognition*, ed. H. Bunke, World Scientific, pp. 333-341, 1992.

[4] S. Latifi, "Correcting OCR-generated Text Using Permutations," *Proc. ICEE 93,* Amir Kabir University, Tehran, Iran, May 1993.

[5] S. V. Rice, "The OCR Experimental Environment, Version 3," University of Nevada, Las Vegas, Technical Report ISRI TR-93-04, April 1993.

[6] S. V. Rice, J. Kanai, and T. A. Nartker, "An Evaluation of OCR Accuracy," University of Nevada, Las Vegas, Technical Report ISRI TR-93-01, April 1993.

[7] J. Kanai, S. V. Rice, and T. A. Nartker, "A Preliminary Evaluation of Automatic Zoning," University of Nevada, Las Vegas, Technical Report ISRI TR-93-02, April 1993.

A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals

Y. S. Huang and C. Y. Suen

***Abstract*—For pattern recognition, when a single classifier cannot provide a decision which is 100 percent correct, multiple classifiers should be able to achieve higher accuracy. This is because group decisions are generally better than any individual's. Based on this concept, a method called the "Behavior-Knowledge Space Method" was developed, which can aggregate the decisions obtained from individual classifiers and derive the best final decisions from the statistical point of view. Experiments on 46,451 samples of unconstrained handwritten numerals have shown that this method achieves very promising performances and outperforms voting, Bayesian, and Dempster-Shafer approaches.**

***Index Items*—Unconstrained handwriting recognition, combination of multiple classifiers, evidence aggregation, behavior-knowledge space, knowledge modeling.**

I. INTRODUCTION

The recognition of handwritten numerals has been studied for more than three decades; during this period, many classifiers with high recognition rates have been developed [1]. However, none of them can achieve satisfactory performance when dealing with characters of degraded quality. A new trend [2], [3], [4], [5], [6] called "Combination of multiple experts" (CME) has emerged to solve this problem. It is based on the idea that classifiers with different methodologies or different features can complement each other. Hence if different classifiers cooperate with each other, group decisions may reduce errors drastically and achieve a higher performance.

In general, based on output information, classifiers can be derived into two types: type-1 outputs a unique class label indicating that this class has the highest probability to which the input pattern belongs; and type-2 assigns each class label a measurement value which indicates the degree that the corresponding class pertains to the input pattern. In fact, type-2 classifiers can be transformed into type-1 ones by outputting only the class with the highest degree. This is an **information reduction** or **abstraction** process. In this sense, all classifiers are type-1 classifiers. Therefore, the research on methods of combining type-1 classifiers becomes most important.

Previous studies have developed many CME approaches of type-1 classifiers, among which the voting [7], Bayesian [5], [8], and Dempster-Shafer (D-S) [5], [9] approaches are the most representative. Simply speaking, voting is a democracy-behavior approach based on "the opinion of the majority wins". It treats classifiers equally without considering their differences in performance. The Bayesian approach uses the Bayesian formula to integrate classifiers' decisions; usually, it requires an independence assumption in order to tackle the computation of the joint probability. The D-S formula, which has frequently been applied to deal with uncertainty management and incomplete reasoning, can aggregate committed, uncommitted and ignorant beliefs. It allows one to attribute belief to subsets, as well as to individual elements of the hypothesis set. Both Bayesian and D-S approaches make use of probability to describe the different qualities of classifiers' decisions. However, in the Bayesian approach, the sum of $P(C)$ and $P(\sim C)$ is equal to one; this is not nec-

Manuscript received October 13, 1993; revised April 25, 1994. Recommended for acceptance by Dr. R. L. Kashyap.

The authors are affiliated with the Centre for Pattern Recognition and Machine Intelligence, Concordia University, Suite GM-606, 1455 de Maisonneuve Blvd. West, Montréal, Québec, H3G 1M8, Canada.

IEEE Log Number P95010.

0162-8828/95$04.00 © 1995 IEEE

essarily true for the D-S approach, where $P(C)$ represents the probability that C is true. Generally speaking, these three approaches, more or less, require the independence assumption which does not usually hold in real applications. This significantly limits their applicability.

Recently, the Behavior-Knowledge Space method (the BKS method) [10] has been developed by the Concordia OCR research team for combining type-1 classifiers. This method offers a number of advantages over previous CME methods. A detailed discussion of this method and two of its properties appears in Section III.

In total, this paper comprises six sections. Section II defines the symbols used and gives a formal description of the CME of type-1 classifiers. Section III introduces the BKS method, and Section IV discusses its two salient properties. Section V describes experiments performed by several CME methods and the comparisons of their results. Finally, Section VI draws our conclusions.

II. Problem Formulation

e_k represents expert (classifier) k where $k = 1,\cdots,K$, and K is the total number of experts. $C_1,\cdots, C_M$ are mutually exclusive and exhaustive sets of patterns. M represents the total number of pattern sets or classes. $\Lambda = \{1,\cdots, M\}$ is a set which consists of all class labels. x is the unknown input pattern and $e_k(x) = j_k$ means expert k assigns the input x to class j_k, where $j_k \in \Lambda \cup \{M+1\}$. When $j_k \in \Lambda$, it means that expert k accepts x (either correct recognition or substitution); otherwise expert k rejects x. To simplify the notation, $e_k(x)$ is replaced by $e(k)$. Then, the research focus on CME becomes, "When K classifiers give their individual decisions $e(1),\cdots, e(K)$ about the identity of x, what is the combination function $E(e(1), \cdots, e(K))$ which can produce the best final decision?"

III. The Behavior-Knowledge Space Method

A thorough analysis of the reason why most CME methods require the independence assumption reveals that either they treat each classifier equally, or they derive information useful for the combination stage from the confusion matrix of a single classifier. Naturally, both situations need to assume that the decisions of classifiers are independent. To avoid this assumption, the information should be derived from a knowledge space which can **concurrently** record the decisions of all classifiers on each learned sample. Since this knowledge space record the behavior of all classifiers, we call it the "Behavior-Knowledge Space". Simply speaking, the BKS method derives its final decisions from a behavior-knowledge space.

A. Behavior-Knowledge Space

A Behavior-Knowledge Space (BKS) is a K-dimensional space where each dimension corresponds to the decision of one classifier. Each classifier has M+1 possible decision values chosen from the set $\{1,\cdots, M+1\}$. The intersection of the decisions of individual classifiers occupies one unit of the BKS, and each unit accumulates the number of incoming samples for each class. The unit which is the intersection of the classifiers' decisions of the current input is called the **focal** unit. Table 1 gives an example of a two-dimensional BKS, where unit (i,j) is the focal unit when $e(1) = i$ and $e(2) = j$. Each unit contains three kinds of data: (1) the total number of incoming samples, (2) the best representative class, and (3) the total number of incoming samples for each class. Learned samples distributed to one unit/class are called incoming units of that unit/class.

TABLE I.
2-D Behavior-Knowledge Space

$e(1) \setminus e(2)$	1	...	j	...	11
1	(1,1)	...	(1,j)	...	(1,11)
⋮	⋮	⋮	⋮	⋮	⋮
i	⋮	⋮	(i,j)	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
11	(11,1)	...	(11,j)	...	(11,11)

Symbols used in defining a BKS are

BKS = a K-dimensional behavior-knowledge space,

BKS($e(1)$, ..., $e(K)$)= a unit of BKS, where classifier 1 gives its decision as $e(1)$, ..., and classifier K gives its decision as $e(K)$,

$n_{e(1)\ldots e(K)}(m)$ = the total number of incoming samples belonging to class m in BKS($e(1)$, ..., $e(K)$),

$T_{e(1)\ldots e(K)}$ = the total number of incoming samples in BKS($e(1)$, ..., $e(K)$),

$$= \sum_{m=1}^{M} n_{e(1)\ldots e(K)}(m), \tag{1}$$

$R_{e(1)\ldots e(K)}$ = the best representative class of BKS($e(1)$, ..., $e(K)$),

$$= \{j \mid n_{e(1)\ldots e(K)}(j) = max_{1\le m\le M} n_{e(1)\ldots e(K)}(m).\} \tag{2}$$

The following is one example of a two-classifiers BKS, which demonstrates the situation where the classifiers give different decisions. Suppose for x, the decision of the first classifier is 4 and that of the second classifier is 9, *i.e.* $e(1) = 4$ and $e(2) = 9$. Obviously, the focal unit now is BKS (4,9). Let there be non-zero incoming samples only for classes 4 and 9 in this unit; and they amount to

$$\begin{cases} n_{49}(4) = 15, \\ n_{49}(9) = 5, \\ n_{49}(\mathrm{m}) = 0, \text{ when } m \notin \{4,\ 9\} \text{and} m \in \Lambda. \end{cases}$$

Then, $T_{49} = 20$ and $R_{49} = 4$.

Interestingly, the semantic meaning of the BKS is clear. For example, in the above, it means when classifier 1 recognizes x to be 4 and classifier 2 to be 9, there is a 75 percent probability that the input x belongs to class 4, and also a 25 percent probability to class 9. Obviously, for any focal unit, if rejection is not allowable, then the class with the highest probability is the best and the safest to choose as the final decision.

B. Two Stage Operations and the Decision Rule

The BKS method operates in two stages: knowledge modeling and decision making. The knowledge-modeling stage uses the learning set of samples with both genuine and recognized class labels to construct a BKS; then the values of $T_{e(1)\ldots e(K)}$ and $R_{e(1)\ldots e(K)}$ of each unit BKS($e(1),\cdots,e(K)$) are computed by equations (1) and (2). The decision-making stage, according to the constructed BKS and the decisions offered from the individual classifiers, enters the focal unit and makes the final decision by the following rule:

$$E(x) = \begin{cases} R_{e(1)\ldots e(k)}, \text{ when } T_{e(1)\ldots e(k)} > 0 \text{ and } \dfrac{n_{e(1)\ldots e(k)}\left(R_{e(1)\ldots e(k)}\right)}{T_{e(1)\ldots e(k)}} \ge \lambda; \\ M+1 \quad \text{, otherwise.} \end{cases} \tag{3}$$

where λ is a threshold ($0 \le \lambda \le 1$) which controls the reliability of the final decision. The knowledge-modeling stage needs to be executed only once for the learning set of patterns, but the decision-making stage will be performed on each test pattern.

IV. Properties

Many good properties can be derived from this method. However, only two of them are presented here due to limited space: automatic threshold finding, and optimality.

A. Automatic Threshold Finding

Among OCR applications, each may have its own specific performance requirements. For example, in monetary applications, the substitution rate should be approximately 0. But, in address-reading applications, slight errors are endurable, such as one error per 1000 pieces of mail. Therefore, it is important that a system should have the capability to automatically adapt itself to the required performance. The following six parameters are used to accomplish this adaptation: (1) $\mathcal{C}$: the required recognition rate, (2) $\mathcal{S}$: the required substitution rate, (3) $\mathcal{R}$: the required rejection rate, (4) DREC(P_t): the derived recognition rate with threshold P_t, (5) DSUB(P_t): the derived substitution rate with threshold P_t, and (6) DREJ(P_t): the derived rejection rate with threshold P_t.

The goal here is to find a threshold P_t for equation (3) which can make the system perform close to the required level. Let $C(P_t)$ be defined as an error function which is the sum of three values, each of which is the square of the distance between the required rate and its derived one, that is

$$C(P_t) = [\mathrm{DREC}(P_t) - \mathcal{C}]^2 + [\mathrm{DSUB}(P_t) - \mathcal{S}]^2 + [\mathrm{DREJ}(P_t) - \mathcal{R}]^2. \quad (4)$$

Since the value of the error function $C(P_t)$ denotes the difference between the desired and derived performances when the threshold is P_t, it is obvious that if the value of $C(P_t)$ is small, then the derived performance is close to the desired one. Therefore, the best P_t (denoted by P^*) is the one which minimizes the value of the error function; that is, $C(P^*) = \min_{0 \le P_t \le 1} C(P_t)$. To derive P^*, some more symbols are defined below :

$$D_{e(1)\ldots e(K)} = \text{the proportion of unit BKS}(e(1), \ldots, e(K)) \text{ to the whole BKS,} \\ = \frac{T_{e(1)\ldots e(K)}}{\sum_{e(1)=1}^{M} \cdots \sum_{e(1)=1}^{M} T_{e(1)\ldots e(K)}},$$

$$P_{e(1)\ldots e(K)} = \text{the probability of the best representative class in unit BKS}(e(1), \ldots, e(K)),$$

and

$$f(P_{e(1)\ldots e(K)}, P_t) = \text{the acceptance index,} \\ = \begin{cases} 1, & \text{when } P_{e(1)\ldots e(k)} \ge P_t, \\ 0, & \text{otherwise.} \end{cases}$$

For a unit BKS($e(1),\cdots,e(K)$), there are two exclusive situations between $P_{e(1)\ldots e(K)}$ and P_t:

(1) $P_{e(1)\ldots e(K)} < P_t$: all samples in this unit are rejected. Therefore, the contribution of this unit to both the derived recognition and substitution rates is 0, and to the derived rejection rate it is $D_{e(1)\ldots e(K)}$.

(2) $P_{e(1)\ldots e(K)} \ge P_t$: all samples in this unit are accepted. Accordingly, $T_{e(1)\ldots e(K)} * P_{e(1)\ldots e(K)}$ samples are recognized correctly and $T_{e(1)\ldots e(K)} * (1 - P_{e(1)\ldots e(K)})$ samples are misrecognized. Therefore, the contribution of this unit to the derived recognition rate is $D_{e(1)\ldots e(K)} * P_{e(1)\ldots e(K)}$, to the derived substitution rate it is $D_{e(1)\ldots e(K)} * (1 - P_{e(1)\ldots e(K)})$, and to the derived rejection rate it is 0.

In fact, the above two situations can be further aggregated together such that the contribution of one unit BKS($e(1)\cdots e(K)$) to the derived recognition rate is $D_{e(1)\ldots e(K)} * P_{e(1)\ldots e(K)} * f(P_{e(1)\ldots e(K)},P_t)$, to the derived substitution rate it is $D_{e(1)\ldots e(K)} * (1 - P_{e(1)\ldots e(K)}) * f(P_{e(1)\ldots e(K)},P_t)$, and to the derived rejection rate it is $D_{e(1)\ldots e(K)} * (1 - f(P_{e(1)\ldots e(K)},P_t))$. Therefore, by summing up the contributions of all units, we get

$$\begin{aligned} \mathrm{DREC}(P_t) &= \sum_{e(1)=1}^{M} \cdots \sum_{e(K)=1}^{M} D_{e(1)\ldots e(K)} * P_{e(1)\ldots e(K)} * f\left(P_{e(1)\ldots e(K)}, P_t\right), \\ \mathrm{DSUB}(P_t) &= \sum_{e(1)=1}^{M} \cdots \sum_{e(K)=1}^{M} D_{e(1)\ldots e(K)} * \left(1 - P_{e(1)\ldots e(K)}\right) * f\left(P_{e(1)\ldots e(K)}, P_t\right), \\ \mathrm{DREJ}(P_t) &= \sum_{e(1)=1}^{M} \cdots \sum_{e(K)=1}^{M} D_{e(1)\ldots e(K)} * \left(1 - f\left(P_{e(1)\ldots e(K)}, P_t\right)\right). \end{aligned}$$

So, the error function $C(P_t)$ becomes

$$\begin{aligned} C(P_t) = & \left[\sum_{e(1)=1}^{M} \cdots \sum_{e(K)=1}^{M} D_{e(1)\ldots e(K)} * P_{e(1)\ldots e(K)} * f\left(P_{e(1)\ldots e(K)}, P_t\right) - \mathcal{C}\right]^2 \\ & + \left[\sum_{e(1)=1}^{M} \cdots \sum_{e(K)=1}^{M} D_{e(1)\ldots e(K)} * \left(1 - P_{e(1)\ldots e(K)}\right) * f\left(P_{e(1)\ldots e(K)}, P_t\right) - \mathcal{S}\right]^2 \\ & + \left[\sum_{e(1)=1}^{M} \cdots \sum_{e(K)=1}^{M} D_{e(1)\ldots e(K)} * \left(1 - f\left(P_{e(1)\ldots e(K)}, P_t\right)\right) - \mathcal{R}\right]^2. \end{aligned}$$

There are many approaches which can find P^*. Accordingly, whenever the required $\mathcal{C}$, $\mathcal{S}$, $\mathcal{R}$ are known, the best threshold P^* can be found from equation (5) automatically.

B. Optimality

From the statistical point of view, after each classifier has given its decision to x, the belief value $BEL(i)$ of x belonging to class i can be computed from a conditional probability as

$$BEL(i) = P(x \in C_i \,/\, e(1) = j_1, \cdots, e(K) = j_K, EN^K)$$

where $P(.)$ is the probability function, and EN^K denotes the classification environment generated from combining the K classifiers. As a matter of fact, (1) EN^K is precisely equal to the BKS of the K classifiers, and (2) the condition under both $e(1)=j_1,\cdots,e(K)=j_K$ and EN^K is actually the same as that of the focal unit BKS $(j_1,\cdots,j_K)$. Based on these two understandings, the belief function becomes

$$BEL(i) = P(x \in C_i \,/\, BKS(j_1, \cdots, j_K)).$$

Using the knowledge of the focal unit BKS $(j_1,\cdots,j_K)$, the belief function can be further modified as

$$BEL(i) = \begin{cases} \dfrac{n_{j_1,\ldots,j_K}(i)}{T_{j_1,\ldots,j_K}}, & \text{when } T_{j_1,\ldots,j_K} > 0; \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

Undoubtedly, the best choice is to assign the unknown input to the class with the highest belief, which is the class with the largest number of incoming samples. With a threshold λ to keep the decision more reliable, the best decision rule finally becomes

$$E(x)=\begin{cases} j & \text{, when } n_{e(1)\ldots e(K)}(j)=\max_{m\in\Lambda} n_{e(1)\ldots e(K)}(m), \\ & T_{e(1)\ldots e(K)}>0, \text{ and } \dfrac{n_{e(1)\ldots e(K)}(j)}{T_{e(1)\ldots e(K)}}\geq\lambda \;; \\ M+1, & \text{otherwise.} \end{cases}$$

Looking at equations (3) and (6), they are essentially identical. Therefore, the performance of the BKS method is equal to that of the optimal one.

V. Experiments

Three classifiers (called e_1, e_2 and e_3) developed by the CENPARMI and ITRI[1] OCR research teams were chosen as three experts. The data used for this experiment come from the ITRI numeral data base, which contains 46,451 numeral samples collected from more than 1,000 people. The whole data was divided into 10 sets. The first set (5,074 samples) was used for training the individual classifiers, and the remaining 9 sets (41,377 samples) for testing their performances.

To index the performances, Rec., Sub., Rej., and Rel. denote the recognition, substitution, rejection, and reliability rates respectively, where the scale of Rec., Sub. and Rej. is percentage. Table II shows the performances of the three classifiers on the 41,377 samples.

TABLE II.
Performance of Individual Classifiers Using 41,377 testing Samples.

	Rec.	Sub.	Rej.	Rel.
e_1	90.37	9.63	0.00	0.9037
e_2	90.93	9.07	0.00	0.9093
e_2	92.14	7.86	0.00	0.9214

The goal of this experiment is to compare the performances of four CME methods: voting, Bayesian, D-S and BKS. For an unbiased comparison, we adopted a leave-one-out estimation [13]. Tables 3(a), 3(b), 3(c) and 3(d) list the results produced by voting, D-S, Bayesian, and BKS approaches, respectively. Fig. 1 shows these results in a graphic form. Two observations may be drawn from this experiment: (1) All four CME methods perform much better than any individual classifier. This shows that by CME a recognition system with high recognition and low substitution rates is achievable. (2) generally, the BKS method performs best among the four CME methods. However, in the situation of a low substitution rate, the BKS method's performance is degraded drastically, and all four approaches get very similar performances. This indicates that if the substitution rate must be very low (such as 0.77 percent), then the simplest voting might be a suitable choice; otherwise, the BKS method is the one to achieve the highest recognitionperformance.

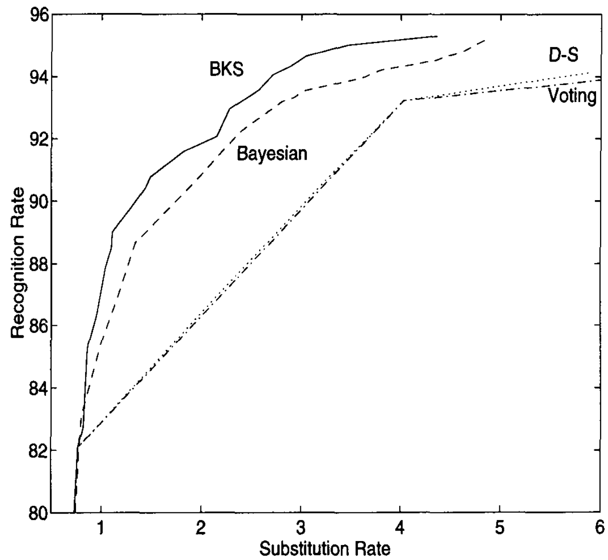


Fig. 1. Graphic representation of the performances of four CME.methods.

VI. Conclusion

Kanal [12] argued that research addressed to the problem of combining multiple classifiers may provide new insight into pattern recognition. Previously, the main efforts focused on the design of one good classifier so that a desired classification rate could be obtained. Now we can also shift our focus. Instead of designing one high performance classifier (the job is usually extremely difficult), we can build a number of different and complementary ones. Each classifier itself may not achieve the desired performance, but the appropriate combination of these individual classifiers may produce a highly reliable performance. In this paper, the authors have presented a CME method (*i.e.* the BKS method) which can efficiently combine type-1 classifiers. Many good properties have been derived from this method, two of them are described in this paper: automatic threshold finding, and optimality. However, all properties exist from the statistical point of view. This means that a large enough and well representative learning data set should exist. If only few samples are collected, or samples are collected randomly and carelessly, the desired properties of this method cannot be guaranteed. Therefore, for practical applications, the key issue to apply this method successfully is to construct a representative training data base. This indicates that more attention should be paid to data collection, which is often ignored in the current research domain.

Acknowledgments

The authors would like to thank Mr. L.T. Tu for providing the recognition results of three individual classifiers.

This work was supported by research grants received from the Natural Sciences and Engineering Research Council of Canada, and the Chinese Character Recognition Project (project no. 35N1100) supported by MOEA, Taiwan, R.O.C.

References

[1] S. Mori, C.Y. Suen, and K. Yamamoto. "Historical Review of OCR Research and Development," *Proceedings of the IEEE*, vol. 80, no. 7, pp. 1029-1057, 1992.

[2] F. Kimura, Z. Chen, and M. Shridhar. "An Integrated Character Recognition Algorithm for Locating and Recognizing Zip Codes," *United States Postal Service - Advanced Technology Conference*, pp. 605-619, Nov. 1990.

[3] R. Bradford and T. Nartker. "Error Correlation in Contemporary OCR Systems," *1st Int. Conf. on Document Analysis and Recognition*, pp. 516-524, 1991.

[4] J. Franke and E. Mandler. "A Comparison of Two Approaches for Combining the Votes of Cooperating Classifiers," *11th International Conference on Pattern Recognition*, vol. 2, pp. 611-614, 1992.

[1] ITRI is a government–sponsored research institution in Taiwan.

TABLE III.
RESULTS FOR 41,377 SAMPLES USING A LEAVE-ONE-OUT ESTIMATION TO COMBINE THE THREE CLASSIFIERS ($e1$, $e2$ AND $e3$).

λ	Rec.	Sub.	Rej.	Rel.
0.000	93.92	6.08	0.00	0.9392
0.333	93.24	4.03	2.03	0.9586
0.667	82.09	0.77	17.14	0.9908

(a). VOTING

λ	Rec.	Sub.	Rej.	Rel.
0.000	94.13	5.87	0.00	0.9413
0.700	93.24	4.03	2.73	0.9586
0.800	90.50	3.20	6.30	0.9659
0.850	87.04	2.18	10.77	0.9755
0.900	82.09	0.77	17.14	0.9908

(b) D-S

λ	Rec.	Sub.	Rej.	Rel.
0.000	95.14	4.86	0.00	0.9514
0.200	94.67	4.52	0.81	0.9544
0.500	94.19	3.82	1.98	0.9610
0.700	93.54	3.05	3.41	0.9684
0.900	92.25	2.49	5.25	0.9737
0.950	90.65	1.97	7.38	0.9787
0.980	88.63	1.34	10.02	0.9851
0.995	85.37	1.03	13.60	0.9881

(c) BAYESIAN

λ	Rec.	Sub.	Rej.	Rel.
0.000	95.31	4.36	0.33	0.9563
0.050	95.03	3.47	1.50	0.9648
0.200	94.08	2.72	3.20	0.9719
0.300	92.97	2.28	4.76	0.9761
0.400	91.60	1.83	6.56	0.9804
0.500	90.42	1.44	8.14	0.9844
0.600	89.01	1.11	9.88	0.9876
0.950	82.09	0.77	17.14	0.9908

(d) BKS

[5] L. Xu, A. Krzyzak, and C.Y. Suen. "Method of Combining Multiple Classifiers and Their Application to Handwritten Numeral Recognition," *IEEE Trans. on Systems, Man and Cybernetics*, vol. SMC-22, no. 3, pp. 418-435, 1992.

[6] T.K. Ho, J.J. Hull, and S.N. Srihari. "Decision Combination in Multiple Classifier Systems," *IEEE Trans. on Pattern Recognition and Artificial Intelligence*, vol. 16, no. 1, pp. 66-75, 1994.

[7] C.Y. Suen, R. Legault, C. Nadal, M. Cheriet, and L. Lam. "Building a New Generation of Handwriting Recognition Systems," *Pattern Recognition Letters*, vol. 14, no. 4, pp. 303-315, 1993.

[8] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Mateo, California, 1988.

[9] M.D. McLeish, P. Yao, and T. Stirtzinger. "A Study on the Use of Belief Functions for Medical Expert Systems," *Journal of Applied Statistics*, vol. 18, no. 1, pp. 155-174, 1991.

[10] Y.S. Huang and C.Y. Suen. "An Optimal Method of Combining Multiple Experts for Handwritten Numerical Recognition," *Pre-Proc. International Workshop on Frontiers in Handwriting Recognition*, pp. 11-20, Buffalo, New York, USA, 1993.

[11] P.A. Devijver and J. Kittler. *Pattern Recognition — A Statistical Approach*, London: Prentice-Hall, 1982.

[12] L.N. Kanal. "On Pattern, Categories, and Alternative Realities," *Pattern Recognition Letters*, vol. 14, no. 3, pp. 241-255, 1993.