A feedback-based multi-classifier system

G. Pirlo [(*)(^)], C.A.Trullo [(*)(^)], D. Impedovo [(§)(^)]

[(*)] *Dipartimento di Informatica – Università degli Studi di Bari – Via Orabona 4–Bari – Italy*
[(§)] *Dipartimento di Elettrotecnica ed Elettronica – Politecnico di Bari– Via Orabona 4–Bari – Italy*
[(^)] *Centro "Rete Puglia" - Università degli Studi di Bari – Via G. Petroni 15/F.1–Bari – Italy*
pirlo@di.uniba.it

Abstract

Multi-classifier approach is a widespread strategy used in many difficult classification problems. Traditionally, in a multi-classifier approach, a classification decision based on the combination of a multitude of classifiers is expected to outperform the decisions of each individual classifier. Therefore, in a multi-classifier systems, the potential of the whole set of classifiers is only exploited at the level of the final decision, in which the contributions of all classifiers is used by combining their individual decisions.

This paper shows a feed-back based multi-classifier system in which the multi-classifier approach is used not only for providing the final decision, but also for improving the performance of the individual classifiers, by means of a closed-loop strategy.

The experimental tests have been carried out in the field of hand-written numeral recognition. The result demonstrates the effectiveness of the proposed approach and its superiority with respect to traditional approach.

1. Introduction

Classifier combination is a widespread strategy to design high-performance classification systems. Many approaches have been proposed so far for classifier combination which can differ in terms of type of output they combine, system topology and degree of a-priori knowledge they use [1,2,3].

Concerning the output they combine, the methods for classifier combination are generally classified into three basic categories: abstract-level combination methods use the top candidate provided by each classifier; ranked-level combination methods use the entire ranked list of candidates; measurement-level combination methods use also the confidence value of each candidate in the ranked list [4, 5].

Concerning system topology, when parallel topology is considered, all the outputs of the classifiers are combined in parallel; when serial topology is adopted, the classifiers are arranged in cascade and the output of each classifier is the input to the classifier of the next stage of the cascade (if any); hybrid topologies combine parallel and serial schemes [4, 5].

Concerning the a-priori knowledge used for classifier combination, some methods do not require any kind of a-priori information on the set of classifiers; other methods need information at the level of each individual classifier; in other cases the combination methods need information on the behavior of the entire set of classifiers [5].

Of course, whatever classifier combination method is considered, the basic consideration is that the performance obtained by combining the outputs of several classifiers of a set can outperform the performance of each classifier of the set. In other words, the main idea of classifier combination is that the collective behavior of a set of classifier can convey more information that those of each classifier of the set, and this information can be exploited for classification aims [6,7,8,9].

On the basis of this consideration this paper addresses the problem to verify the possibility that each individual classifier of the set can learn from the collective behavior of the whole set. For this purpose, a feed-back based parallel topology is considered and some experimental tests are carried out in the field of classification of handwritten numerals. Moreover, two combination rules are considered: the Majority Vote and the Sum Rule [1,2,4].

The experimental results demonstrate that, in some cases, it is possible to use the output of a multi-expert system to improve the performance of the individual classifiers and, finally, to improve the performance of the whole system.

The paper is organized ad follows: Section 2 presents the new feed-back based parallel topology; Section 3 shows the process of learning of the individual classifiers based on the collective behavior of the multi-expert system; the experimental results are presented in Section 4; Section 5 presents some discussions and the conclusion of the work.

IEEE computer society

2. A feed-back based topology

In a traditional multi-classifier system using a parallel topology (see Fig. 1a), the input pattern x_t is fed to K individual classifiers in parallel. Each classifier A_i provides a response $A_i(x_t)$, on the basis of the information stored in its knowledge base KB_i, $i=1,2.,..K$. The responses obtained by all the classifiers are then combined to obtain the final results $E(x_t)$ according to a suitable combination strategy $E(A_1(x_t), A_2(x_t),..., A_K(x_t)) \rightarrow E(x_t)$ [1,4].

(a)

(b)
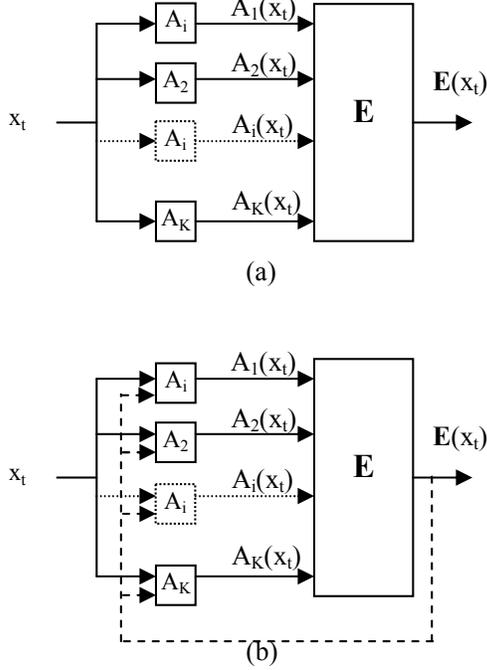
Figure 1. Multi-classifier Parallel Systems.

In the *feed-back* based *parallel* topology here proposed (see Fig. 1b), the final results $E(x_t)$ provided by the multi-expert system is used for updating the knowledge base of those individual classifiers whose decisions disagree with $E(x_t)$. For instance, in the case of *abstract-level combination*, if $E(x_t) \neq A_i(x_t)$ then *update KB_i*, $i=1,2,...K$. Conversely, in the case of *measurement-level combination*, KB_i, $i=1,2,...K$, should be updated if $E(x_t) \neq A_i^T(x_t)$, being $A_i^T(x_k)$ the top candidate provided by A_i.

3. Learning from collective behavior

In a feedback-based multi-expert approach, the learning process and the classification process are here considered as a unique, dynamical process, as it happens for human beings. In this process, after an initial training stage, the multi-expert system performs at the same time classification and learning, when necessary. More precisely, let be:

- A_i, for i=1,2,..., N, the set *abstract level* classifiers;
- C_j, for j=1,2,..., M, the set of pattern classes;
- $P=\{x_k \mid j=1,2,...,K\}$ the set of patterns, that is here considered as partitioned into S subsets $P_1, P_2, ..., P_s, ..., P_S$, being
  - $P_s=\{x_k \in P \mid k \in [N_s \cdot (s-1)+1, N_s \cdot s]\}$ and $N_s=K/S$ (N_s integer), that are fed one after the other to the multi-expert system.

  In particular, P_1 is used for learning only, whereas $P_2, P_3,...,P_s,...,P_S$ are used both for classification and learning (when necessary);

- $F_i(k) = (F_{i,1}(k), F_{i,2}(k), ..., F_{i,r}(k),... F_{i,R}(k))$ the numeral feature vector used by A_i for representing the pattern $x_k \in P$ (for the sake of simplicity it is here assumed that each classifier uses R numeral features).

Furthermore, let be $KB_i(s) = (KB_i^1(s), KB_i^2(s), ..., KB_i^j(s),..., KB_i^M(s))$ (s=1,2,...,S), the knowledge base of A_i after the subsets $P_1, P_2, ...,P_s$ have been processed, being $KB_i^j(s)$ the component of the knowledge base concerning the class j, j=1,2,...,M.

Initially, at the first stage (s=1), the classifier A_i is trained using the patterns $x_k \in P_i^* = P_1$. Therefore, the knowledge base $KB_i(s)$ of A_i is initially defined as:

$$KB_i(s) = (KB_i^1(s), KB_i^2(s),...,KB_i^j(s),...,KB_i^M(s)) \quad (1a)$$

where, for j=1,2,...,M:

$$KB_i^j(s)=(F_{i,1}^j(s), F_{i,2}^j(s),..., F_{i,r}^j(s),..., F_{i,R}^j(s)) \quad (1b)$$

being $F_{i,r}^j(s)$ the average value of the r-th feature of the i-th classifier for the patterns of the class C_j that belongs to P_i^*, i.e. :

$$F_{i,r}^j(s) = \frac{1}{Card(C_j(P_i^*))} \cdot \sum_{x_k \in C_j(P_i^*)} F_{i,r}(k) \quad ,r=1,2,,R \quad (1c)$$

with $C_j(P_i^*) = \{x_k \in C_i \mid x_k \in P_i^*\}$ (it is here assumed that $C_j(P*_i) \neq \varnothing$, $\forall j=1,2,...,M$).

Successively, the subsets $P_2, P_3,...,P_s,...,P_S$ are provided one after the other to the multi-classifier system both for classification and for learning. In particular, after the subsets $P_2, P_3,...,P_s$ have been fed into the system, the knowledge base $KB_i(s)$ of A_i, for i=1,2,...N, is updated according to eqs. (1) and using the patterns $x_k \in P_i^*$, with:

$$P_i^* = P_1 \cup \{x_k \in P_2 \cup P_3 \cup... \cup P_s \mid A_i^T(x_k) \neq E(x_k)\},$$

where $A^T_i(x_k)$ is the top candidate provided by A_i (when measurement–level combination is considered). In other words, each input pattern, for which the response provided by the multi-classifier system is different from that generated from an individual classifier A_i, is used to update the knowledge base of A_i.

4. Experimental Results

In order to evaluate experimentally the proposed approach, a multi-expert system for handwritten numeral recognition has been considered. It uses the following individual classifiers whose details can be found in refs. [10, 11, 12, 13] : A_1: Histograms; A_2: Intersections; A_3: Zoning; A_4: Pattern Matching. Moreover, the following methods have been used for decision combination [1,4,14,15]: C_1: Majority Vote (MV), for *abstract-level combination*; C_2: Sum Rule (SR), for *measurement-level combination* .

	A_1	A_2	A_3	A_4	MV	SR
"0"	84%	96%	80%	99%	95%	96%
"1"	36%	99%	82%	84%	87%	88%
"2"	58%	52%	63%	37%	58%	60%
"3"	65%	67%	73%	70%	77%	82%
"4"	74%	60%	52%	13%	54%	66%
"5"	37%	10%	84%	47%	35%	33%
"6"	55%	28%	81%	30%	57%	61%
"7"	51%	51%	52%	43%	48%	55%
"8"	40%	42%	37%	58%	44%	60%
"9"	17%	57%	72%	69%	64%	75%

Table 1. Performance.

A database $P=\{x_k \mid j=1,2,\ldots,1300\}$ of 1300 numerals (classes from "0" to "9") extracted from the CEDAR database has been used [16]. Furthermore, it has been partitioned into 10 subsets:

- $P_1=\{x_1,x_2,x_3,\ldots, ,x_{130}\}$,
- $P_2=\{x_{131},x_{132},x_{133},\ldots, ,x_{260}\}$,
- …,
- $P_j=\{x_{1+130*(j-1)},x_{2+130*(j-1)},x_{3+130*(j-1)},\ldots, ,x_{130+130*(j-1)}\}$,
- …,
- $P_{10}=\{x_{1+130*(10-1)},x_{2+130*(10-1)},x_{3+130*(10-1)},\ldots, ,x_{1300}\}$.

Table 1 reports the performance of the individual classifiers and of the combination methods, estimated when the patterns in P_1 and in $P_2 \cup P_3 \cup\ldots , \cup P_{10}$ are used for learning and testing, respectively.

In the first set of tests, in order to verify the extent to which the results depend on the order in which the subsets are arranged, different subset sequence I-J have been considered, being I the index of the initial subset and J the skip used to determine the other subsets in the sequence. In other words a Subset Sequence I-J is defined as $[P_{(I+(n-1)\cdot J)\bmod 10}]_{n=1,2,\ldots,10}$. In particular, in the experimental tests, four subset sequences have been considered according to a random criterion: a) sequence 1-1; b) sequence 4-3; c) sequence 5-1; d) sequence 2-3. Thus, when the Subset Sequence 1-1 is considered, the subset P_1 is used to train individually each classifier. Successively, the other subsets are provided consecutively to the system, according to the natural order P_2, $P_3,\ldots,P_{10}$. Conversely, when the Subset Sequence 4-3, 5-1 and 2-3 are considered, the subsets P_4 , P_5 and P_2 are used – respectively - to train individually each classifier. Successively, the other subsets are provided consecutively to the system, according to the following orderings P_7, P_{10}, P_3, P_6, P_9, P_2, P_5, P_8, P_1; P_6, P_7, P_8, P_9, P_{10}, P_1, P_2, P_3, P_4 and P_5, P_8, P_1, P_4, P_7, P_{10}, P_3, P_6, P_9, respectively.

Figure 2 shows comparatively the recognition rates of the system using the feed-back based procedure, with respect to the traditional parallel systems, in which no feedback was activated. As it can be noted, benefits can be registered whatever sequence is considered for inputting the data subsets. For instance, when MV and SR are considered, the feed-back based system outperforms the traditional system respectively in the 130% and 20% of the cases, if the sequence 1-1 is considered; in the 40% and 20% of the cases, if the sequence 4-3 is considered; in the 75% and 50% of the cases, if the sequence 5-1 is considered. Finally, when the sequence 2-3 is considered, the feed-back based system outperforms the traditional system only if MV is used. In this case the improvement occurs in the 40% of the cases. Conversely, no improvement can be registered in using the feed-back based system, if SR is used; this result merits special consideration because it impose to understand better the cases under which learning can really improve system performances.

Furthermore, on average, when MV is considered, the feed-back based system outperforms the traditional system in the 71% of the cases; conversely, when SR is used, the feed-back based system provides superior performance with respect to the traditional system in the 23% of the cases.
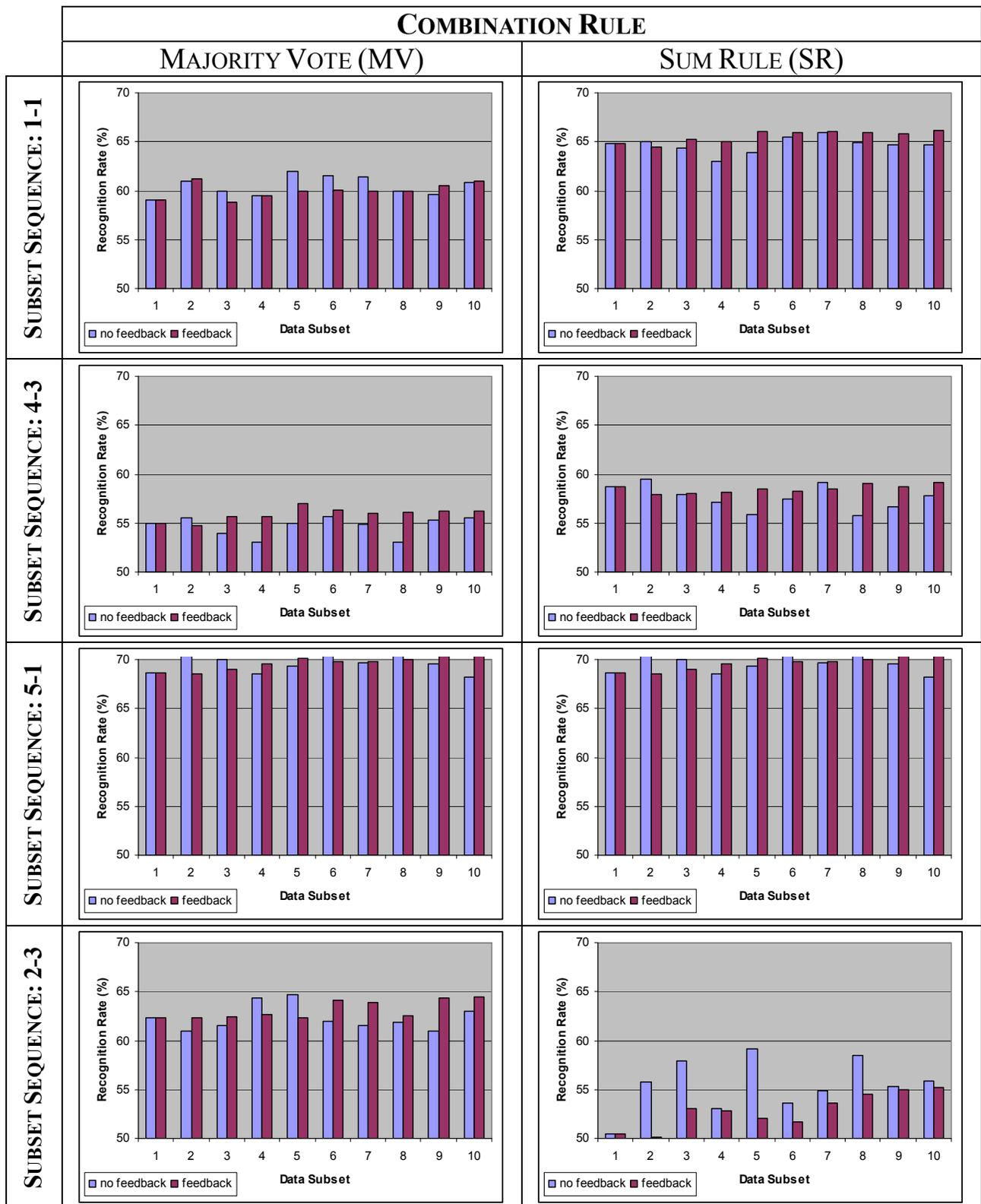
FIGURE 2. Performances.

5. Discussion and Conclusion

This paper shows the possibility to improve the effectiveness of a multi-classifier system by a suitable use of the information extracted from the collective behaviour of the classifiers. More precisely, the final decision obtained by combining the individual decisions provided by each classifier, has been used to upgrade the knowledge base of the individual classifiers, when necessary, according to a feed-back based topology. The experimental results reported in this paper demonstrate that the collective behaviour of a set of classifier *can* provide useful information to improve system performance to a certain extent.

Of course, the definition of the most profitable procedure to upgrade the individual knowledge-base of each classifier is not trivial and poses many problems that need additional research. Specifically, it is mainly related to the definition of *when* and *how* the knowledge base of each classifier should be upgraded, on the basis of the behaviour of the individual classifier itself and of the collective behaviour of the entire set of classifiers. For instance, concerning *when* the knowledge base of each classifier should be upgraded, it could be argued that the upgrading of a knowledge-base of a classifier can be a complex, time consuming task. Therefore it should be performed exclusively in the case in which it can significantly improve system performance, according to the specific requirements of the application. In fact, it should be considered the risk that, in an extremely unfortunate scenario, the output of the combined system is wrong and it can drag down the performance of the most valuable classifiers. Concerning *how* the knowledge base of each classifier should be upgraded, it should be considered that, although the upgrading of the knowledge base of the individual classifiers can lead to an improvement of the performance of each individual classifiers, it can also lead to a reduced complementarity in the individual behaviours, reducing the overall performance of the multi-classifier system.

Finally, it should also be investigated the extent to which the effectiveness of a feed-back based strategy depends on the specific technique used for decision combination.

In conclusion, this paper shows the performance of a multi-classifier system can be improved by exploiting the collective behaviour of the set of classifiers. Of course, in order to make this strategy feasible, additional research is necessary.

References

[1] J. Kittler, M. Hatef, R.P.W. Duin, J. Matias, "On combining classifiers", *IEEE Trans. on Pattern Analysis Machine Intelligence*, Vol.20, no.3, pp.226-239, 1998.

[2] R. Plamondon, S.N. Srihari, "On-line and Off-line Handwriting Recognition: A comprehensive survey", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, n.1, pp. 63-84, 2000.

[3] C.Y.Suen, C. Nadal , R.Legault, T.A.Mai, L.Lam, "Computer Recognition of unconstrained handwritten numerals", *Proc. IEEE*, Vol. 80, pp. 1162-1180, 1992.

[4] Ley Xu, Adam Krzyzak, Ching Y-Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition" , *IEEE Transaction on Systems, Man and Cybernetics*- Vol. 22, N. 3, 1992, pp. 418-435.

[5] V. Di Lecce, G.Dimauro, A. Guerriero, S.Impedovo, G.Pirlo, A. Salzo, "Classifier Combination: the role of a-priori knowledge", *Proc. of IWFHR-7*, Sept. 2000, Amsterdam, The Netherlands, pp. 143-152.

[6] *C.Y. Suen, J. Tan,* Analysis of errors of handwritten digits made by a multitude of classifiers", *Pattern Recognition Letters*, No. 3, Feb. 2005, pp. 369-379.

[7] G.Dimauro, S.Impedovo, G.Pirlo, "Multiple Experts:A New Methodology for the Evaluation of the Combination Processes", IWFHR-5, Colchester,Uk,1996,pp.131-136.

[8] S. Impedovo and A. Salzo, "A new methodology for expert combination in multi-expert system designing". In *Lecture Notes in Computer Science*, (vol. 1857), J. Kittler and F.Roli Eds., MCS2000, Cagliari, Italy, pp.230-239.

[9] Lucchese M.G., Impedovo, S. Pirlo, G., "Optimal Zoning Design by Genetic Algorithms", *IEEE Trans. Sys. Men and Cybern. - Part A*, Vol. 36 , Issue: 5, Sept. 2006, pp. 833-846.

[10] G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo, "Automatic Bankchecks Processing: A New Engineered System", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.11, N.4, World Scientific Publ., Singapore, 1997, pp.1-38.

[11] S. Mori, C.Y.Suen, K.Yamamoto, "Historical Review of OCR research and development", *Proc. IEEE*, Vol. 80, pp. 1029-1058, 1992.

[12] O.D.Trier, A.K.Jain , T.Taxt, "Feature Extraction Methods For Character Recognition – A Survey", *Pattern Recognition*, Vol. 29, n.4, pp. 641-662, 1996.

[13] M. Bokser , "Omnidocument Technologies", *Proc. IEEE* , Vol. 80, 1992, pp. 1066-1078.

[14] C.Y.Suen, J.Guo, Z.C.Li , "Analysis and Recognition of Alphanumeric Handprints by Parts", *IEEE Trans. Systems, Man and Cybernetics*, Vol. 24, n. 4, pp. 614-630, 1994.

[15] F. Kimura , M. Shridhar, "Handwritten Numerical Recognition Based on Multiple Algorithms", *Pattern Recognition*, Vol. 24, n. 10, pp. 969-983, 1991.

[16] J. Hull, "A database for handwritten text recognition research", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, n. 5, pp. 550–554, 1994.