A Multi-Objective Memetic Algorithm for Intelligent Feature Extraction

Paulo V. W. Radtke[1,2], Tony Wong[1], and Robert Sabourin[1,2]

[1] École de Technologie Supérieure
Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielle,
Department de Génie de la Production Automatisé,
1100, rue Notre Dame Ouest, Montréal, QC, Canada H3C 1K3

[2] Pontifícia Universidade Católica do Paraná
Rua Imaculada Conceição 1155, Curitiba, PR, Brazil 80215-901
radtke@livia.etsmtl.ca**

Abstract. This paper presents a methodology to generate representations for isolated handwritten symbols, modeled as a multi-objective optimization problem. We detail the methodology, coding domain knowledge into a genetic based representation. With the help of a model on the domain of handwritten digits, we verify the problematic issues and propose a hybrid optimization algorithm, adapted to needs of this problem. A set of tests validates the optimization algorithm and parameter settings in the model's context. The results are encouraging, as the optimized solutions outperform the human expert approach on a known problem.

1 Introduction

Image-based pattern recognition (PR) systems require that pixel information be first transformed into an abstract representation suitable for recognition, a process called feature extraction [1]. A methodology that extracts features for PR must select the most appropriate transformations and determine the spatial location of their application on the image. Related to the feature extraction process is the feature subset selection (FSS) operation [2]. FSS further refines the extraction process by selecting the most relevant features, within the extracted feature set, in order to reduce classifier's computation effort in the classification stage and improve recognition rate. A comparison of FSS methods in [3] indicates that genetic algorithm (GA) based approach performs better than traditional methods when the problem size is large (more than 50 features). In the context of isolated handwritten digits, Oliveira et al. applied a GA based FSS [4] to optimize classifier accuracy and feature set cardinality using a weighted vector. They postulated that a multi-objective genetic algorithm (MOGA) could further enhance the obtained results. Their postulate was later confirmed in [5], where

** Author for correspondence.

MOGA outperformed GA on the same problem. The superiority of MOGA in FSS is also confirmed by Emmanouilidis et al. using sonar and ionosphere data [6].

It is now understood that the advantage of MOGA lies in the inherent diversity of the optimized solution set, avoiding the population convergence to a single local optimum. However, the application of MOGA in FSS also faces a number of difficulties. In essence, classifier training is based on a finite set of labeled observations – the training set. The classifier may perform differently when presented to unknown observations, i.e., data not in the training set. This behavior is verified in [5] where the feature set optimized by a MOGA produces a recognition rate that is different for the optimization set compared to a set of unknown observations. This behavior can be explained by the fact that the input domain used in the MOGA optimization process does not match the one used in the classification stage of the recognition process. Thus, the corresponding objective spaces are also non matching because the same classifier is used in the optimization and the classification stages. Another difficulty arises when two or more feature sets sharing similar elements exist in the MOGA population. In the context of FSS, similar feature sets should yield comparable performances for a given classifier. If these feature sets also possess the same cardinality then the genetic selection operator is likely to emphasize the one with the highest recognition rate. Since the FSS problem has non matching objective spaces, the selected feature set may not perform adequately in the classification of unknown observations. Furthermore, the genetic selection operation is complicated by the dominance principle used in most Pareto-based MOGAs. The primary aim of the FSS operation is to reduce feature set cardinality while maintaining the highest possible recognition rate. This implies a mixed-integer objective space and standard dominance relationship can not be implemented directly. Due to the existence of the L_1 norm, special steps must be taken in order to ensure diversity on the Pareto-front. Finally, due to the non matching input domains and objective spaces, a non dominated feature set in the optimization stage is not necessarily non dominated in the classification stage.

Considering these aspects, we propose in Sect. 2 a methodology for feature extraction of isolated handwritten symbols formulated as an evolutionary multi-objective optimization problem (MOOP), supported by earlier experiments in [7]. Section 3 analyze the MOOP and verify the issues discussed in the context of isolated handwritten digits. Sections 4 to 6 describe an optimization algorithm adapted to the FSS problem and present a series of tests to verify its efficiency. Section 7 presents the conclusions.

2 The Intelligent Feature Extractor Methodology

Traditionally, human experts are responsible for the choice of the feature set. It is most often determined by using domain knowledge on a trial and error basis. We propose to use the domain knowledge in a methodology formulated as an MOOP to genetically evolve a set of candidate solutions – the Intelligent Feature

Extractor (IFE) methodology. The goal of this work is to help the human expert in defining representations (feature sets) in the context of isolated handwritten symbols.

2.1 IFE Concepts

The IFE methodology models handwritten symbols as features extracted from specific *foci* of attention on images using *zoning*. It is a strategy known to provide better results in recognition than features extracted from the whole image [8]. In the proposed IFE three operators are needed to generate representations: a *zoning operator* to define *foci* of attention over images, a *feature extraction* operator to apply transformations in zones, and a *feature subset selection* operator that removes irrelevant features. The domain knowledge introduced by the human expert lies in the choice of transformations for the feature extraction operator.

The operators are combined to generate a representation, as illustrated by Fig. 1. The *zoning operator* defines the zoning strategy $Z = \{z^1, \ldots, z^n\}$, where $z^i, 1 \leq i \leq n$ is a zone in the image I and n the number of zones. The pixels inside the zones in Z are transformed by the *feature extraction* operator in the representation $F = \{f^1, \ldots, f^n\}$, where f^i is the feature vector extracted from z^i. F has the irrelevant features eliminated by the *feature subset selection operator*, producing the representation $G = \{g^1, \ldots, g^n\}$, g^i being the feature subset of f^i.

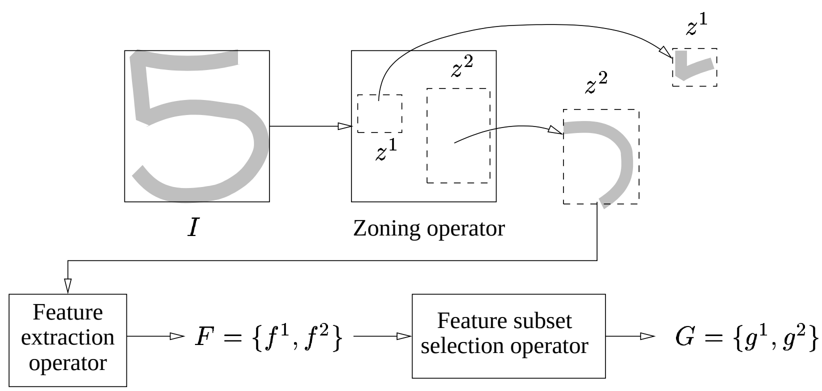


Fig. 1. IFE hierarchical structure

Candidate solutions are represented using a hierarchical genetic coding, with three different parts. Each part of the genetic coding is related to an IFE operator, as shown in Fig. 2. The parts are hierarchical in the sense that the coding in one part will determine the data manipulated by another. After optimization the result is a set of representations. The human expert can either select the representation with the highest accuracy, or use the result set to optimize an ensemble of classifiers (EoC) [9] for improved accuracy.

Zoning	*Feature extraction operators*	*Feature subset selection*

Fig. 2. IFE candidate solution coding

2.2 Dividers Zoning Operator

To compare the IFE against the traditional human expert approach we consider a baseline representation known to achieve high accuracy with isolated handwritten digits [10]. The zoning on this representation can be defined as a set of three dividers, where the intersection of image borders and dividers defines zones as 4-sided polygons. Here we expand this concept into a set of 5 horizontal and 5 vertical dividers that can be either *active* or *inactive*. Figure 3 details the operator template, represented by a 10 bits binary string, each bit associated to a divider. This operator produces zoning strategies with 1 to 36 zones, and the baseline zoning in [10] can be obtained by setting d_2, d_6 and d_8 active.

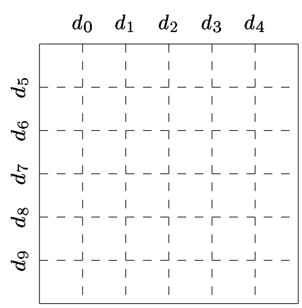


Fig. 3. Dividers zoning operator

2.3 Feature Extraction Operator

In [10], Oliveira et al. used a mixture of concavities, contour and surface transformations, extracting 22 features per zone – 13 for concavities, 8 for contour and 1 for surface. We consider that the performance achieved on handwritten digits supports the use of these transformations to optimize representations. With three different transformations the operator is encoded as a three bits binary string, where each bit indicates the state of the associated transformation. When all transformations are inactive, the zone becomes a missing part [11], a zone with no features extracted.

2.4 Feature Subset Selection Operator

The feature subset selection operator selects the most relevant features in the feature vector $F = \{f^1, \ldots, f^n\}$, creating a final representation $G = \{g^1, \ldots, g^n\}$.

This task is performed with a binary string associated to each feature set f^i, where each bit indicates if the associated feature in f^i is active or not. Thus a 22 bits binary string is required to encode the feature extraction operator described in the previous section.

3 IFE Model

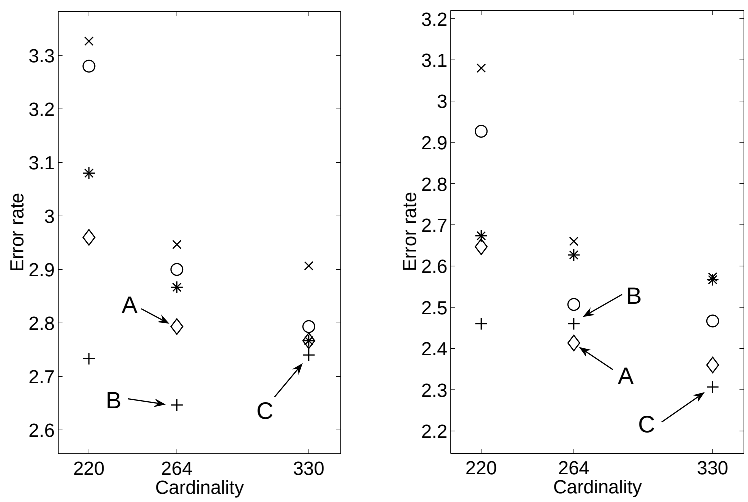
To verify the issues discussed in Sect. 1 we created a model, based on the dividers zoning operator with all 22 features extracted from each zone. In this model, only the dividers zoning operator is active, while the feature extraction and feature subset selection operators have fixed values – to extract all features from all zones. We calculated the entire objective space, evaluating the representations discriminative power on a *wrapper* approach [2] using actual classifier performance. We used the projection distance (PD) classifier [12], with the databases of digits in Table 1 to calculate representations error rate. The disjoint databases are extracted from the NIST SD19 database, a widely used database of isolated handwritten symbols, using the digits data sets *hsf_0123* and *hsf_7*. In this objective space, we minimize both the feature set cardinality and the error rate.

To train the PD classifier we use the *learn* database as the learning examples, and the *validation* database to configure classifier parameters during learning. The error rate during the IFE optimization is calculated with the trained classifier on the *optimization* database. In order to verify the generalization power of solutions optimized by the IFE, we use the *selection* and *test* databases to compare the error rates on unknown observations.

Table 1. Handwritten digits databases

Database	Size	Origin	Sample range
learn	50000	hsf_0123	1 to 50000
validation	15000	hsf_0123	150001 to 165000
optimization	15000	hsf_0123	165001 to 180000
selection	15000	hsf_0123	180001 to 195000
test	60089	hsf_7	1 to 60089

Figure 4 partially details the error rates of solutions using the model's precalculated objective space. Solutions A and C are dominated by solution B on the *optimization* database objective space – Fig. 4.a. In this context, solution B belong to the Pareto-optimal set. Solutions A and B have the same cardinality, but the later has lower error rate, and solution B outperforms solution C in both feature set cardinality and performance. Changing the context to the *selection* database to evaluate the error rate in Fig. 4.b, we have that solution B is dominated by solution A, and that solution C becomes non-dominated. Both solutions A and C belong to the Pareto-optimal set in the *selection* database context.



(a) *optimization* database (b) *selection* database

Fig. 4. Objective space – PD classifier

Because of the non matching objective function spaces, it is clear that the IFE needs a post-processing stage to analyze and select the optimized solutions regarding the generalization power on unknown observations. Therefore we define the IFE optimization in two stages. The first stage optimizes solutions on the *optimization* database objective space. The second stage analyzes solutions archived by the optimization algorithm in the *selection* database objective space, where the IFE user selects one or more solutions, based on the error rate. This yields two requirements that must be satisfied by an MOGA algorithm for proper optimization of the IFE methodology:

1. Optimize the best solution for each cardinality value, which we call the *decision frontier* of the objective space.
2. Archive different levels of performance regarding solutions cardinality.

4 Multi-Objective Memetic Algorithm

The Multi-Objective Memetic Algorithm (MOMA) combines a traditional MOGA with a local search (LS) algorithm, featuring modified selection and archiving strategies suitable for the IFE methodology. The combination of MOGA with LS is discussed in [13], and [14] demonstrates that hybrid methods outperform methods solely based on genetic optimization in some problems.

4.1 Concepts

To store the decision frontier defined in Sect. 3, we divide the objective functions in two categories, *objective function one* (o_1) in the integer domain, that defines the *slots* of our archive, and *objective function two* (o_2), which is optimized for each o_1 value. To archive different levels of performance, the slot S^l is a set of max_{S^l} solutions, associated to a possible value of o_1. For our IFE problem, o_1 is the feature set cardinality and o_2 is the error rate.

The archive is defined as $S = \{S^1, \ldots, S^j\}$, where j is the maximum number of slots. For solution X^i, $o_1(X^i)$ and $o_2(X^i)$ are the solution's values of o_1 and o_2. $B(S^l) = \{X^i \in S^l | o_2(X^i) = min(o_2(x)), \forall x \in S^l\}$ indicates the solution X^i in S^l with the best o_2 value, while $W(S^l) = \{X^i \in S^l | o_2(X) = max(o_2(x)), \forall x \in S^l\}$ indicates the opposite.

The decision frontier set P_S optimized by the MOMA algorithm is defined as $P_S = \bigcup_{l=1}^{j} \{B(S^l)\}$. We indicate that solution X^i is admissible into slot S^l as $X^i \bowtie S^l \equiv o_1(X^i) = o_1(x \in S^l)$, then $A(S^l, C) = \{c | c \in C \wedge c \bowtie S^l\}$ denotes the subset of solutions in C that are admissible in S^l.

To optimize the decision frontier, solutions are ranked for genetic selection by a *frontier ranking* approach. In the population P, the solution set belonging to the first rank is defined by $R^1 = \bigcup_{l=1}^{j} \{B(A(S^l, P))\}$. The solution set belonging to the second rank R^2 is obtained as the first rank of $P \backslash R^1$, and so on.

The decision frontier concept and the archive S are key elements for proper optimization of the IFE methodology. Combined they provide means to select solutions after optimization based on their generalization power on unknown observations. Selecting solutions by the decision frontier allows the optimization of solutions usually discarded by traditional Pareto based approaches. This need is justified to avoid optimization bias as indicated in Fig. 4, where solution C is dominated in the IFE model optimization objective space using the *optimization* database, but has better generalization power on unknown observations. The same principle justifies the need to store different levels of performance in the slot, solution A in Fig. 4 has better generalization power on unknown observations, but would be discarded from the archive on traditional approaches.

4.2 Algorithm Discussion

The MOMA algorithm is depicted in Fig. 5. It evolves a population P of size m, and archives good solutions found in the slots S, which are updated at the end of each generation. The population P is initialized in two steps. The first creates candidate solutions with a Bernoulli distribution, while the second generates individuals to initialize the slots. For each slot, we choose one random solution that is admissible in the slot and insert it in the population.

During genetic optimization, individuals in the current generation P_t are subjected to frontier ranking. Next a mating pool M is created by tournament selection, followed by crossover and mutation to create the offspring population

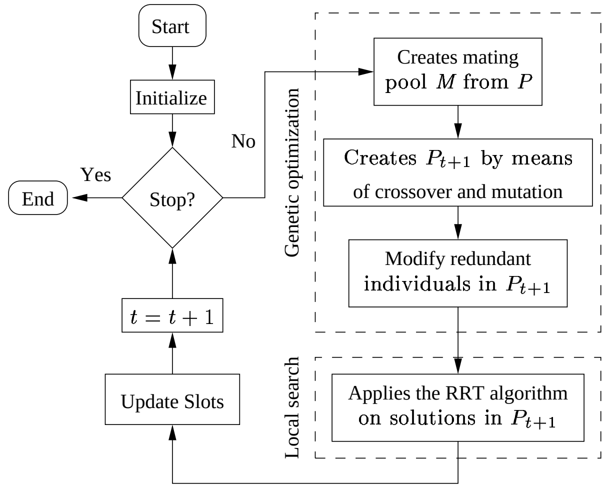


Fig. 5. MOMA algorithm

P_{t+1}. In case of a draw in the tournament selection, one of the solutions is chosen randomly. To avoid genetic overtake, redundant individuals are mutated until the population has no redundant individuals as in [9].

After genetic optimization solutions are further improved by a LS algorithm. We choose the *Record-to-Record Travel* (RRT) algorithm [15], an annealing based heuristic. The RRT algorithm improves solutions by searching in its neighborhood for n potential solutions during NI iterations, allowing a decrease in the current performance of $a\%$ to avoid local optimal solutions.

Neighbors to solution X^i must have the same feature set cardinality and similar structure, which is achieved in the IFE model by modifying the zoning operator encoding. The model defined in Sect. 3 has all features extracted from all zones, and the feature extraction and feature subset selection operators are fixed. With the zoning operator dividers distributed in two groups, $g_1 = \{d_0, d_1, d_2, d_3, d_4\}$ and $g_2 = \{d_5, d_6, d_7, d_8, d_9\}$, to generate a neighbor we select a group to activate one divider and deactivate another. The solution in Fig. 6.a has solutions in Figs. 6.b and 6.c as two possible neighbors.

After the LS, the archive S is updated, storing good solutions from P_{t+1} in the slots as in Algorithm 1. Recall that max_{S^l} is the maximum number of solutions a slot can hold. At this point, we verify the stopping criterion, deciding if the algorithm should continue to the next iteration or stop the optimization process.

5 Experimental Protocol

To test the MOMA algorithm, we conducted three tests on the IFE model's objective space, optimizing only the IFE zoning operator, while keeping the

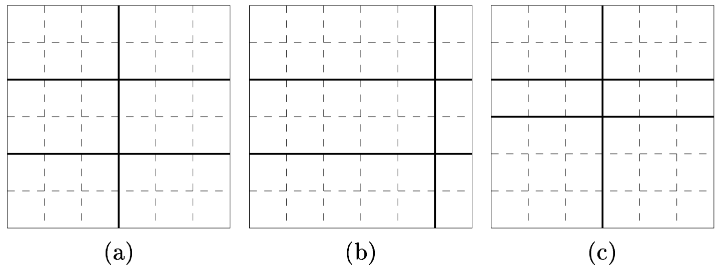


Fig. 6. Solution (a) and two neighbors $(b$ and $c)$

forall $x^i \in P$ **do**
  Determines the slot S^l solution x^i relates to;
  if $E(x^i) < E(W(S^l))$ *and* $x^i \notin S^l$ **then**
    $S^l = S^l \cup \{x^i\}$;
    if $|S^l| > max_{S^l}$ **then**
      $S^l = S^l \backslash \{W(S^l)\}$;
    end
  end
end

Algorithm 1: Update slot algorithm

remaining operators fixed to extract all 22 features from each zone. All tests used the *learn* and *validation* databases to train the PD classifier, and the *optimization* database to evaluate the error rate during the optimization process. The first test verifies that the genetic optimization has convergence properties in this type of problem. We achieve this by disabling the RRT algorithm with $NI = 0$. The second test evaluates the MOMA algorithm with a neighborhood subset best improvement strategy, while the third test uses a greedy first improvement strategy, where $n = 1$ and $a = 0\%$. We call these tests as *Test A*, *Test B* and *Test C*, respectively.

The usual genetic operators for the IFE are the *hierarchical single point crossover*, where the single point crossover is performed independently over each IFE operator in the chromosome with probability p_c, and the *hierarchical bitwise mutation*, which also performs the bitwise mutation independently over each IFE operator in the chromosome with probability $p_m = 1/L$, where L is the length in bits of the encoded operator being mutated. For the IFE model, genetic operations are restricted to the zoning operator, in order to extract all features from each zone.

We defined a set of values for each algorithmic parameter, using a *fractional design* approach [16] to obtain the 18 configuration sets in Table 2. During *Test A* and *Test C*, we replace columns in this table with specific values to achieve the desired effects.

Each configuration set is subjected to 30 runs of 500 generations in each test, and comparisons are made at the generation of convergence. One run is said to

Table 2. Parameter values

#	m	p_c	p_m	a	n	NI
1	32	70%	$1/L$	5%	2	7
2	32	70%	$1/L$	5%	3	5
3	32	70%	$1/L$	5%	4	3
4	32	80%	$1/L$	4%	2	7
5	32	80%	$1/L$	4%	3	5
6	32	80%	$1/L$	4%	4	3
7	32	90%	$1/L$	2%	2	7
8	32	90%	$1/L$	2%	3	5
9	32	90%	$1/L$	2%	4	3
10	64	70%	$1/L$	5%	2	7
11	64	70%	$1/L$	5%	3	5
12	64	70%	$1/L$	5%	4	3
13	64	80%	$1/L$	4%	2	7
14	64	80%	$1/L$	4%	3	5
15	64	80%	$1/L$	4%	4	3
16	64	90%	$1/L$	2%	2	7
17	64	90%	$1/L$	2%	3	5
18	64	90%	$1/L$	2%	4	3

have converged when the optimized decision frontier set P_S can no longer be improved. Preliminary experiments indicated that 500 generations far exceed the number of generations required to converge *Test A*, our worst case scenario. Thus the following metrics are used to compare runs:

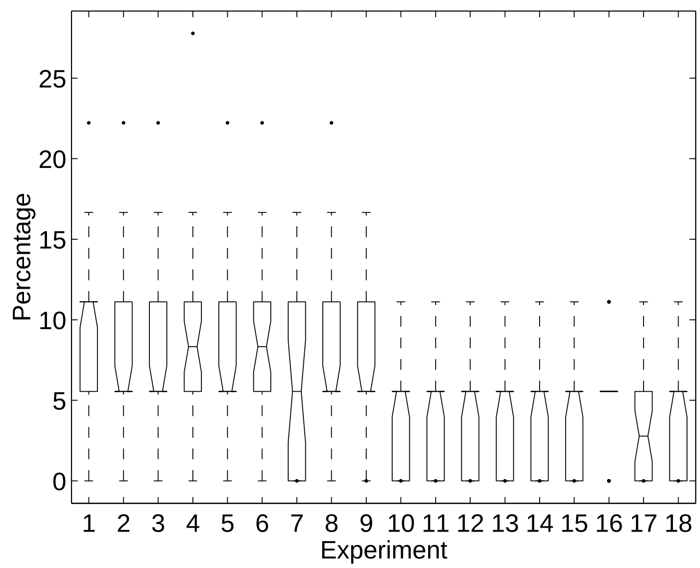
1. *Unique individual evaluations* – how many unique individuals have been evaluated until the algorithm convergence, which relates to the computational effort.
2. *Coverage by the global optimal set* – percentage of individuals in P_S that are covered by solutions in the global optimal set [17], adapted to the decision frontier context. When P_S converges to the optimal set, the coverage is equal to zero.

Both metrics are fair as they hold the same meaning for all three tests. A final test evaluates representations optimized by the MOMA algorithm in the IFE model. We select a result set S, evaluate the error rate of these solutions with the *selection* database and calculate the decision frontier P_S. From this decision frontier we select a set of solutions for testing to compare with the baseline representation.

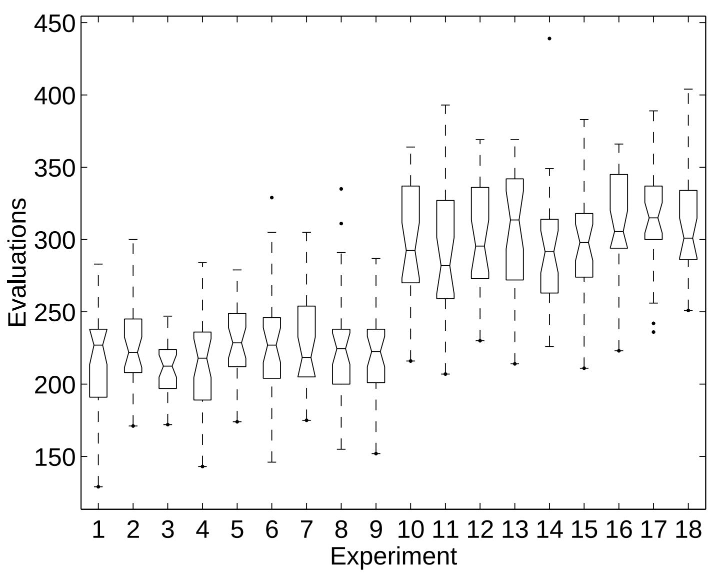
6 Results

The results for the MOMA tests are presented in Figs. 7 to 9. The horizontal axis on the plots relate to configuration sets in Table 2. Experiments 1 to 9 represent a smaller population – 32 individuals, while experiments 10 to 18 represent a larger population – 64 individuals. The box plots summarize the values attained in the 30 runs of each configuration set.

The results for *Test A* in Fig. 7.a indicate the convergence property of genetic operations alone, which is capable to optimize an approximation to the global optimal. The best coverage values where achieved by the larger population, which also explored better the objective space. The exploratory aspect is measured as the number of unique individual evaluations in Fig. 7.b.



(a) Coverage – *Test A*



(b) Unique individual evaluations – *Test A*

Fig. 7. Results – *Test A*

To improve convergence and the objective space exploration, we use in *Test B* the complete MOMA algorithm. The RRT algorithm improved convergence and all runs but a few outliers in the smaller population converged to the optimal set. Objective space exploration in *Test B* is improved, as the number of unique individual evaluations in Fig. 8 is higher than in *Test A* – Fig. 7.b. This improvement reflects in the convergence toward the global optimal set, which is better than in *Test A*. The LS helps to improve convergence when searching for better solutions, which may also helps the genetic algorithm to better explore the objective space.

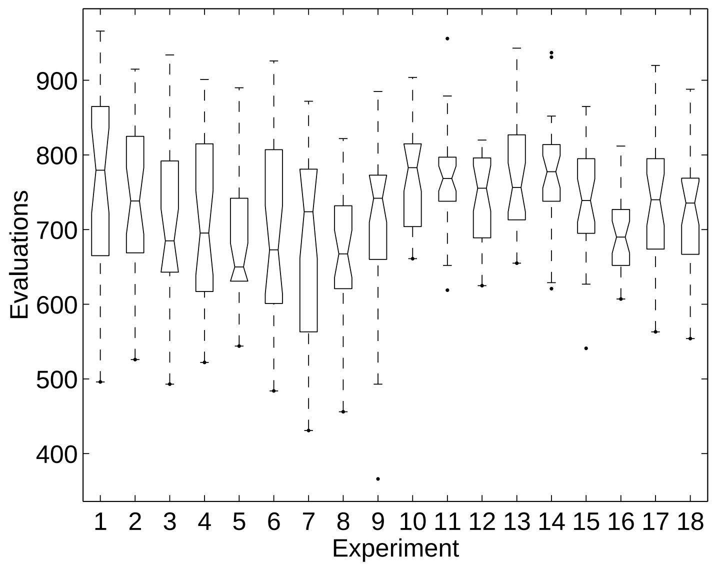


Fig. 8. Unique individual evaluations – *Test B*

In the IFE we are concerned with the error rate evaluation cost. Thus, it is desirable to restrain the number of unique individual evaluations by reducing the strength of the LS. *Test C* modifies the RRT algorithm behavior, using a greedy first improvement strategy. The convergence is similar to *Test B* – all runs but a few outliers converged to the optimal set. However, the number of unique individual evaluations in Fig. 9 is lower than in Fig. 8, which suggests that this improvement strategy is more suitable for the IFE problem optimization.

These results demonstrate the effectiveness of the MOMA algorithm with the IFE methodology, reaching solutions that traditional MOGA approaches could not. For better convergence with lower number of unique individual evaluations, the LS with the greedy first improvement strategy is most appropriate. As for

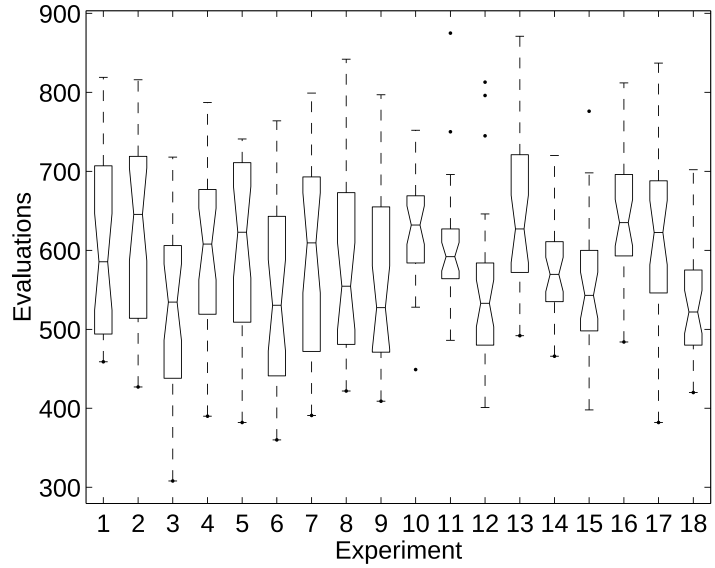


Fig. 9. Unique individual evaluations – *Test C*

the configuration parameters, configuration set 15 in Table 2 modified for the greedy improvement strategy ($n = 1$ and $a = 0\%$) is a good trade-off between convergence and number of unique individual evaluations.

Our final test evaluates a set of solutions optimized by the MOMA algorithm in the IFE model. We selected a random run from *Test C* and evaluated the error rate of solutions in S with the *selection* database. Then we arbitrarily selected solutions a to g from the decision frontier P_S calculated with the error rates on the *selection* database, as discussed in Sect. 3. Finally we tested the selected solutions with the *test* database to compare with the baseline representation. The results are presented in Table 3, where the baseline representation was also trained and evaluated with the PD classifier using the same database set. The table details the feature set cardinality, the binary string associated to the zoning operator and the error rate in three databases, *optimization*, *selection* and *test* – e_{opt}, e_{sel} and e_{test}, respectively.

The results in Table 3 demonstrate that the IFE methodology is able to optimize and select solutions that outperform the traditional human expert approach in the domain of unknown observations – the *test* database. Representation g's error rate is 26.33% lower than the baseline on the *test* database, which justifies the IFE methodology for actual applications. As the model is a subset of the complete methodology, future experiments with the complete IFE methodology are expected to achieve at least this performance level.

Table 3. Representations comparison

Representation	features	zoning operator	e_{opt}	e_{sel}	e_{test}
Baseline	132	00100 01010	3.527%	3.010%	2.959%
a	110	00000 01111	3.587%	3.053%	3.272%
b	132	00000 11111	3.260%	2.987%	2.930%
c	176	00010 01101	3.153%	2.980%	2.981%
d	198	01010 01101	3.273%	2.993%	2.521%
e	220	00100 01111	2.733%	2.460%	2.438%
f	264	01100 01110	2.647%	2.460%	2.568%
g	330	00110 01111	2.740%	2.307%	2.180%

7 Conclusions

This paper presented and assessed the IFE methodology in the context of a model, generating representations for isolated handwritten digits that outperform the human expert approach. These representations where optimized with the proposed MOMA algorithm, an hybrid MOGA approach adapted to the objective space of the IFE problem.

The IFE model demonstrated that the objective space during optimization and the objective space with a set of unknown observations are non matching, which is also verified in the literature on MOGA based FSS. We attribute this to the supervised learning stage, based on a finite set of examples, hence the same behavior is expected in similar optimization problems.

8 Acknowledgments

The first author would like to acknowledge CAPES and the Brazilian government for supporting this research through scholarship grant BEX 2234-03/3.

References

1. Heutte, L., Paquet, T., Moreau, J.V., Lecourtier, Y., Olivier, C.: A structural/statistical feature based vector for handwritten character recognition. Pattern Recognition Letters **19(7)** (1998) 629–641.
2. John, G. H., Kohavi, R., Pfleger, K.: Irrelevant Features and the Subset Selection Problem. Proceedings of the International Conference on Machine Learning (1994) 121–129.
3. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. Pattern Recognition **33(1)** (2000) 25–41.
4. Oliveira, L. S., Benahmed, N., Sabourin, R., Bortolozzi, F., Suen, C. Y.: Feature Subset Selection Using Genetic Algorithms for Handwritten Digit Recognition. Proceedings of the XIV Brazilian Symposium on Computer Graphics and Image Processing (2001) 362–369.

5. Oliveira, L. S., Sabourin, R., Bortolozzi, F., Suen, C. Y.: Feature Subset Selection Using Multi-Objective Genetic Algorithms for Handwritten Digit Recognition. Proceedings of the 16th International Conference on Pattern Recognition **I** (2002) 568–571.
6. Emmanouilidis, C., Hunter A., MacIntyre J.: A Multiobjective Evolutionary Setting for Feature Selection and a Commonality-Based Crossover Operator. Proceedings of the 2000 Congress on Evolutionary Computation (2000) 309–316.
7. Radtke, Paulo V. W., Oliveira, L. S., Sabourin, R., Wong, T.: Intelligent Zoning Design Using Multi-Objective Evolutionary Algorithms. Proceedings of the 7th International Conference on Document Analysis and Recognition (2003) 824–828.
8. Sabourin, R., Genest, G., Prêteux, F. J.: Off-Line Signature Verification by Local Granulometric Size Distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence **19(9)** (1997) 976–988.
9. Ruta, D., Gabrys, B.: Classifier Selection for Majority Voting. Accepted to Information fusion (2004).
10. Oliveira, L. S., Sabourin, R., Bortolozzi, F., Suen, C. Y.: Automatic Recognition of Handwritten Numerical Strings: A Recognition and Verification Strategy. IEEE Transactions on Pattern Analysis and Machine Intelligence **24(11)** (2002) 1438–1454.
11. Li, Z.-C., Suen, C. Y.: The partition-combination method for recognition of handwritten characters. Pattern Recognition Letters **21(9)** (2000) 701–720.
12. Kimura, F., Inoue, S., Wakabayashi, T., Tsuruoka, S., Miyake, Y.: Handwritten Numeral Recognition using Autoassociative Neural Networks. Proceedings of the International Conference on Pattern Recognition **1** (1998) 152–155.
13. Knowles, J.D., Corne, D. W.: Memetic Algorithms for Multiobjective Optimization: Issues, Methods and Prospects. In Recent Advances in Memetic Algorithms. Krasnogor, N., Smith, J.E., and Hart, W.E. (eds). pp. 313-352, Springer, 2004.
14. Jaszkiewicz, A.: Do Multiple-Objective Metaheuristics Deliver on Their Promise? A Computational Experiment on the Set-Covering Problem. IEE Transactions on Evolutionary Computation **7(2)** (2003) 133-143.
15. Pepper, J., Golden, B. L., Wasil, E. A.: Solving the Traveling Salesman Problem With Annealing-Based Heuristics: A Computational Study. IEEE Transactions on Systems, Mand and Cybernetics – Part A: Systems and Humans **32(1)** (2002) 72–77.
16. Gunst, R. F., Mason, R. L.: How to Construct Fractional Factorial Experiments – ASQC basic references on quality control: v. 14. American Society for Quality Control – Statistics Division, 611 East Wisconsin Avenue, Milwaukee, Wisconsin 53202, USA (2001).
17. Zitzler, E., Thiele, L.: Multiobjective Optimization Using Evolutionary Algorithms – A Comparative Case Study. Parallel Problem Solving from Nature – PPSVN V, Springer Verlag (1998).