

Approximation of Digital Curves using a Multi-Objective Genetic Algorithm

Hervé Locteau, Romain Raveaux, Sébastien Adam, Yves Lecourtier, Pierre Héroux, Eric Trupin
LITIS Labs – University of Rouen, FRANCE
Herve.Locteau@univ-rouen.fr

Abstract

In this paper, a digital planar curve approximation method based on a multi-objective genetic algorithm is proposed. In this method, the optimization/exploration algorithm locates breakpoints on the digital curve by minimizing simultaneously the number of breakpoints and the approximation error. Using such an approach, the algorithm proposes a set of solutions at its end. The user may choose his own solution according to its objective. The proposed approach is evaluated on curves issued from the literature and compared successfully with many classical approaches.

1. Introduction

Approximation of digital planar curves using vertices and/or circular arcs is an important issue in pattern recognition and image processing. It is a classical way to represent, store and process digital curves. For example, approximation results are frequently used for shape recognition. The problem can be stated as follows: Given a curve $C = \{C_i \equiv (x_i, y_i)\}_{i=1}^N$ constituted of N ordered points, the goal is to find a subset $S = \{S_i \equiv (x_i, y_i)\}_{i=1}^M$ of M ordered points and the corresponding parameter set $P = \{P_i \equiv (xc_i, yc_i)\}_{i=1}^M$.

S contains the extremities of the line segments or the circular arcs (sometimes called breakpoints) and P the parameters of the best approximation of the set of points between each couple of breakpoints (a specific value is applied in the case of segment)

Whereas many paradigms have been proposed to solve the problem of polygonal approximation or the problem of approximation with circular arcs, much less papers were proposed concerning the approximation of digital curves with both representations. Among the existing papers [1][2][3][4], an approach recently proposed in [4] consists in using Genetic Algorithms

(GA) in order to find a near-optimal approximation. In such a case, the approximation of digital curves is considered as an optimization process. The algorithm automatically selects the best points of the curves by minimizing a given criterion. In [4], the number N of breakpoints to be obtained is fixed and the method uses the concept of genetic evolution to obtain a near-optimal approximation.

In this paper, we adopt the same paradigm and we propose a new GA for the approximation of digital curves. The originality of the described approach is the use of a multi-objective optimization process. Such a new viewpoint enables the user of the system to choose a trade-off between different quality criteria after a single run of the GA.

The remainder of the paper is organized as follows. In section 2, an introduction to the multi-objective optimization problem is proposed and our algorithm is presented. In section 3, the application of this algorithm to the approximation problem is shown. Section 4 presents the experimentally obtained results and a comparison with existing approaches. Section 5 gives the concluding remarks.

2. Multi objective optimization GA

When an optimization problem involves more than one objective function, the task of finding one or more optimum solutions is known as multi-objective optimization. Some classical textbooks on this subject have been published, e.g. [5]. We just recall here some essential notions in order to introduce the proposed algorithm. The main difference between single and multi-optimization tasks lies in the requirement of compromises between the various objectives in the multi-optimization case. Even with only two objectives, if they are conflicting, the improvement of one of them leads to a deterioration of the other one. For example, in the context of polygonal approximation, the decrease of the approximation error always leads to an increase of the vertices number. Two main approaches are used to overcome this

problem in the literature. The first one consists in the combination of the different objectives into a single one (the simpler way being to use a linear combination of the various objectives), and then to use one of the well-known techniques of single objective optimization (like gradient based methods, simulated annealing or classical genetic algorithm). In such a case, the compromise between the objectives is a priori determined through the choice of the combination rule. The main critic addressed to this approach is the difficulty to choose a priori the compromise. It seems a better idea to postpone this choice after having several candidate solutions at hand. This is the goal of Pareto based method using the notion of dominance between candidate solutions. A solution dominates another one if it is better for all the objectives. This dominance concept is illustrated on figure 1. Two criteria J_1 and J_2 have to be minimized. The set of non-dominated points that constitutes the Pareto-Front appears as 'O' on the figure, while dominated solutions are drawn as 'X'.

Using such a dominance concept, the objective of the optimization algorithm becomes to determine the Pareto front, that is to say the set of non-dominated points. Among the optimization methods that can be used for such a task, genetic algorithms are well-suited because they work on a population of candidate solutions. They have been extensively used in such a context. The most common algorithms are VEGA – Vector Evaluated Genetic Algorithm – [6], MOGA – Multi Objective Genetic Algorithm – approach [7], NSGA – Non-Dominated Sorting Genetic Algorithm – [8], NSGA II [9], PAES – Pareto Archived Evolution Strategy – [10] and SPEA – Strength Pareto Evolutionary Algorithm – [11]. The strategies used in these contributions are different, but the obtained results mainly vary from the convergence speed point of view. A good review can be found in [12].

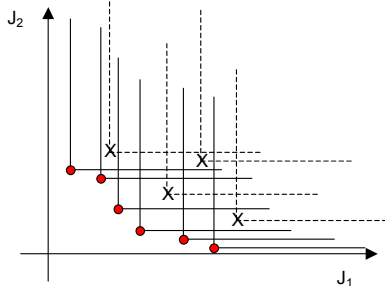


Fig. 1. Illustration of the Pareto Front concept

The proposed genetic algorithm is elitist and steady-state. This means that (i) it manages two populations and (ii) the replacement strategy of individuals in the populations is not made as a whole, but individual per individual. The two populations are a classical

population, composed of evolving individuals and an “archive” population composed of the current Pareto Front elements. These two populations are mixed during the iterations of the genetic algorithm. The first population guarantees space exploration while the archive guarantees the exploitation of acquired knowledge and the convergence of the algorithm.

Based on such concepts, our optimization method uses the following algorithm:

Population (I) and Archive (A) Initialization
do

- Random selection of two individuals I_1 and I_2 in (I)
- Crossover between I_1 and I_2 to generate I_3 and I_4
- Mutation applied to the generated children I_3 and I_4
- Evaluation of children I_3 and I_4
- Selection either of the dominant individual I_5
 between mutated children (if it exists) or random selection of I_5 between I_3 and I_4
- Random selection of (I_6) a in (A)
- Crossover between I_5 and I_6 to generate I_7 and I_8
- Evaluation of children I_7 and I_8
- Test for the integration of I_7 and I_8 in (A)
- Test for the integration of I_7 and I_8 in (I)

While $i <$ the maximal generation number

This algorithm has been designed in order to be applied to various problems. The design of a new application consists in the choice of a coding scheme for individuals, in the design of the evaluation method and in the choice of both parameters values and of some specific operators. In its current implementation, the coding of an individual is a classical bit string. Crossover is a well-known 2-points crossover whereas initialization and mutation are application-dependent. Concerning the replacement strategy, several choices can be made for the integration of a candidate individual in the archive. The simplest is a dominance test between the candidate and the archive elements. The candidate is inserted within the archive if no archive element dominates it. In the same time, archive elements dominated by the candidate are eliminated from the archive. A problem reported in the literature on evolutionary multi-objective optimization is the possible bad exploration of Pareto front: the archive population elements concentrate on only some parts of the front. This difficulty is overcome in our approach by defining a minimal distance between two points in the objective space. This algorithm has been tested on classical multi-objective problems such as BNH or TNK [13]. The obtained results have shown the quality of the proposed approach since it is able to find a similar approximation of the Pareto Front for the same number of calls to the evaluation function.

3. Application to curve approximation

In order to apply the algorithm presented above to the curve approximation problem, an individual has to represent a possible solution to the approximation problem. That is why an individual is composed of N genes, where N is the number of points in the initial curve. A gene is set to '1' if the point is kept as a breakpoint, '0' if it is not. An example of an individual coding is given in figure 2. Each point C_i of the curve S corresponds to a bit in the chromosome. In the example of figure 2, the individual is a binary string of 45 genes corresponding to the initial C_1-C_{45} . The approximation is composed of 2 line-segments and 6 circular arcs. The corresponding breakpoints are respectively C_3 , C_5 , C_{20} , C_{29} , C_{35} , C_{37} , C_{41} , and C_{44} . Such an approximation (the optimal approximation for 8 breakpoints) corresponds to the individual "0010100000000000001000000010000010100010010".

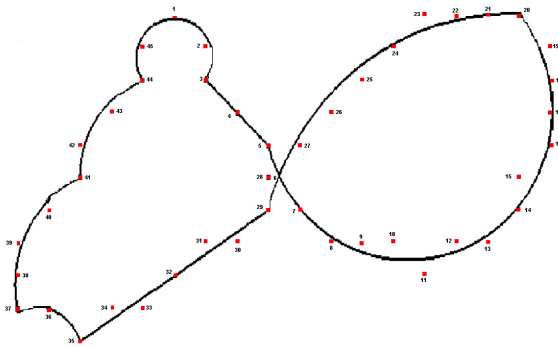


Fig. 2: An example of the coding scheme

Using such a coding scheme, the GA described in section 2 is applied. In order to reduce the number of iteration in the GA, a specific initialization operator is used. It is based on a simple analysis of the curve to be approximated. An histogram of the curvature along the curve is first computed. During initialization, for each point, a probability to be selected is deduced from this histogram. This strategy enables to avoid the selection of collinear points and on the contrary enables to select points with high curvature. A specific mutation operator is also used. It is based on the shift of a selected point to the preceding or the next one. Concerning the criteria to be optimized, two objectives have been included in the current version. The first one is the Integral Square Error (ISE) and the second one is the number of points. This enables to have a trade-off between the precision of the result and the number of line segments, thanks to elements of the Pareto front. One can note that the use of a discrete objective (vertices number) guarantees itself the diversity on the

Pareto front, we do not need to specify any minimal distance between any couples of solutions of the Pareto Front. For the computation of the ISE, the error is computed both in the case of line-segments and circular arcs and the best solution is kept as P_i . Circular arcs are obtained using a LMS approach [4].

4. Experimental results

In order to assess the performances of the proposed algorithm, it has been applied to the four broadly used digital curves presented in [14] and proposed in Fig. 3.

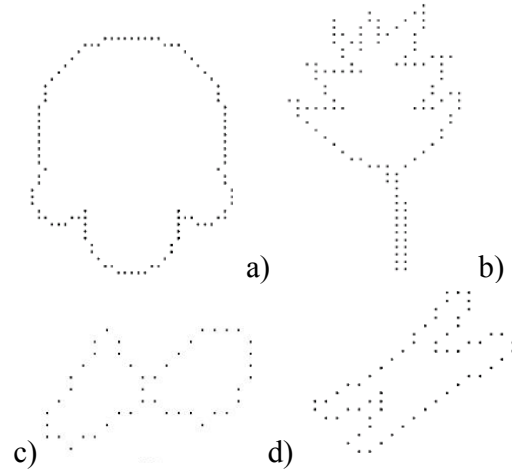


Fig. 3. The four digital test curves

Such tests allow to test the performances of the proposed algorithm versus those of published approaches. For each of these curves, the program has been run for 2000 generations, using a population size of 100 individuals. Such a parameter set involves about 8000 calls to the evaluation method (see the algorithm below). The mutation rate has been fixed to 0.05 and the crossover rate to 0.6. As said before, the output of the presented algorithm is not a single ISE for a number of vertices given a priori. It consists in the whole Pareto front of the optimization problem. That is why the result is a set of couple (ISE – number of vertices). As an example, figure 4 shows the set of couple obtained at the end of the algorithm applied on the "semicircle" curve. Another remark has to be done. Since GA are stochastic, results may be different at independent runs. That is why, in these experiments, we give (table 1) both the best (B) and the worst (W) ISE for each number of vertices obtained after 5 independent runs on each curve. The obtained results can be compared with the results of table 2 issued from an existing comparative study [4]. As one can see on table 1 and 2, results obtained using the GA approach

enables to obtain competitive results. Moreover, these tables also show the stability of the proposed approach since best (B) and worst (W) results are generally the same for the 5 runs.

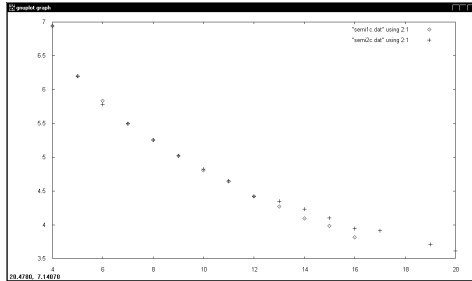


Fig. 4: Two obtained approximations

Table 1 : Results obtained using the GA

| Fig 3a | | | Fig 3b | | | Fig 3c | | | Fig 3d | | |
|--------|-----|-----|--------|------|------|--------|-----|-----|--------|-----|-----|
| N | B | W | N | B | W | N | B | W | N | B | W |
| 4 | 6.9 | 6.9 | 12 | 43.5 | 43.9 | 5 | 5.2 | 5.9 | 9 | 4.6 | 4.6 |
| 5 | 6.1 | 6.1 | 14 | 22.1 | 22.7 | 6 | 3.0 | 3.4 | 10 | 2.4 | 2.4 |
| 6 | 5.7 | 5.8 | 16 | 10.7 | 10.7 | 7 | 2.6 | 2.8 | 11 | 1.9 | 2.0 |
| 7 | 5.4 | 5.7 | 18 | 7.3 | 7.4 | 8 | 2.3 | 2.3 | 12 | 1.6 | 1.7 |
| 8 | 5.2 | 5.2 | 25 | 3.2 | 3.3 | 9 | 1.9 | 1.9 | 13 | 1.4 | 1.4 |
| 12 | 4.2 | 4.4 | 27 | 2.9 | 3.0 | 10 | 1.5 | 1.5 | 14 | 1.2 | 1.2 |
| 14 | 3.8 | 4.0 | 29 | 2.7 | 2.8 | 11 | 1.2 | 1.3 | 15 | 1.0 | 1.1 |
| 22 | 2.3 | 2.4 | 31 | 2.6 | 2.8 | 13 | 0.7 | 0.8 | 16 | 0.9 | 1.0 |

Table 2 : Best results found in the literature for the approximation of the curves of figure 3

| Fig 3a | | Fig 3b | | Fig 3c | | Fig 3d | |
|--------|------|--------|------|--------|-----|--------|-----|
| N° | ISE | N° | ISE | N° | ISE | N° | ISE |
| 4 | 6.9 | 16 | 10.9 | 6 | 3.0 | 10 | 2.6 |
| 6 | 6.4 | 18 | 7.4 | 8 | 2.3 | 11 | 2.1 |
| 12 | 10.9 | 27 | 8.8 | 9 | 2.0 | 15 | 1.2 |
| 14 | 17.7 | 29 | 14.9 | 13 | 5.9 | | |
| 22 | 20.6 | 31 | 1.6 | | | | |

5. Conclusion and future works

In this paper, we have proposed a new approach for the approximation of curves. This approach is inspired from previous approaches in the way that it considers the polygonal approximation as an optimization process. The fundamental difference with the existing approaches lies in the fact that we use a multi-objective optimization process while other contributions only optimize a unique objective, that is to say the ISE. One can see several interests in such an approach. As many solutions are proposed, the user or the system may choose the optimal solution regarding its constraints. Another interest is to offer the possibility to add a new objective easily. As an example, such an approach may be used for the vectorization of shape contours by adding a parallelism constraint.

7. References

- [1] C. Ichoku, B. Deffontaines and J. Chorowicz, "Segmentation of digital plane curves: a dynamic focusing approach", *Pattern Recognition Letters*, 17, 1996, pp 741-750.
- [2] P.L. Rosin and G.A.W. West, "Nonparametric segmentation of curves into various representations", *IEEE Trans. Pattern Anal. Machine Intell.*, 17, 1995, pp 1140-1153.
- [3] J-H. Horng and J.T. Li, "A dynamic programming approach for fitting digital planar curves with line segments and circular arcs", *Pattern Recognition Letters*, 22, 2001, pp 183-197.
- [4] B. Sarkar, L.K. Singh and D. Sarkar, "Approximation of digital curves with line segments and circular arcs using genetic algorithms", *Pattern Recognition Lett.* 24, 2003, 2585-2595.
- [5] K. Deb, "Multi-Objective optimization using Evolutionary algorithms", Wiley, London, 2001.
- [6] J.D. Schaffer and J.J. Grefenstette, "Multiobjective learning via genetic algorithms", *In Proceedings of the 9th international joint conference on artificial intelligence*, Los Angeles, California, pp 593-595, 1985.
- [7] C.M. Fonseca, P.J. Fleming, "Genetic algorithm for multi-objective optimization: formulation, discussion and generalization", In Stephanie editor, *Proceedings of the fifth international conference on genetic algorithm*, San Mateo, California, pp 416-423, 1993.
- [8] N. Srinivas, K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithm", *Evolutionary Computation* 2, 1994, pp 221-248.
- [9] K. Deb, S. Agrawal, A. Pratab and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation* 6, 2000, pp 182-197.
- [10] J.D. Knowles, D.W. Corne, "Approximating the nondominated front using the Pareto archived evolution strategy", *Evolutionary computation* 8, 2000, pp 149-172.
- [11] E. Zitzler, L. Thiele, "Multiobjective evolutionary algorithms : a comparative study and the strength pareto approach", *IEEE Transactions on Evolutionary Computation* 3, 1999, pp 257-271.
- [12] C. A. Coello Coello, "A short tutorial on Evolutionary Multiobjective Optimisation", In Eckart Zitzler, Kalyanmoy Deb, Lothar Thiele, Carlos A. Coello Coello and David Corne (editors), *First International Conference on Evolutionary Multi-Criterion Optimization*, Lecture Notes in Computer Science, . Springer-Verlag n° 1993, pp 21-40, 2001.
- [13] D. Chafekar, J. Xuan, K. Rasheed, "Constrained Multi-objective Optimization Using Steady State Genetic Algorithms", *In Proceedings of Genetic and Evolutionary Computation Conference*, Chicago, Illinois, pp 813-824, 2003.
- [14] R.T. Teh and Chin, "On the detection of dominant points on digital curves", *IEEE transaction on Pattern Analysis and Machine Intelligence* 23 , 1989, pp 859-872.