An Algorithm for Fairing Data Points Based on Genetic Algorithm

Zizhi Lin
College of Mathematics and computer Science
Jiangxi science & Technology Normal University line
Nanchang, China
Linzizhi161@163.com

Sihui Shu
College of Mathematics and computer Science
Jiangxi science & Technology Normal University line
Nanchang, China
sihuishu@163.com

Ding Yun
Science and Technology College
Jiangxi Normal University line 3: City, Country
Nanchang, China
8580090@qq.com

Mincheng Hu
College of Mathematics and computer Science
Jiangxi science & Technology Normal University line
Nanchang, China
humincheng@163.com

Abstract—An algorithm for faring data points is presented in the paper. In this algorithm, researchers split data point set into several smaller point sets according to the number of data point set and features points of the point set, then genetic algorithm will be applied to fair each small point set. Details of this genetic such as fitness function to evaluate the fairness of point set, selection, crossover and mutation are also described in this paper. In order to accelerate its convergence speed, researchers develop the detection algorithm of feature points and the point set subdivision algorithm to control the size of population in per iteration. Experimental results show the efficiency of this method.

Keywords-Component; Genetic Algorithm; Data Point Set ; Fairin

I. INTRODUCTION

In the industrial product design and manufacturing process, people require the product with more smooth appearance. Fair lines and smooth products will lead to higher level of attention; at the same time, smooth products are a great convenience for the processing procedure, and natural transition means less damage to the machine. Data point set fairing is a core issue in reverse engineering and it is also a crucial step [1,2] for curve and surface reconstruction. In the CAD/CAM, data points from the scanner often have errors, and then it is very difficult to construct a smooth curve/surface to fit these data points. Therefore, it is necessary to fair data points before constructing curve or surface from noise data points.

Data points fairing methods have been widely studied in [1,3,4]. In general, if there is a smooth curve passing through these data points, researchers believe that these data points are smooth. There are two methods for faring the data point set: whole fair method and local fair method [4]. As for the whole fair method, the most commonly used methods are minimizing energy of the fitting curve [1] and the least square method [4]. Energy method is used to measure the smoothness of the curve through the strain energy of this curve; the goal for smoothing is to minimize weighted sum of the energy and total deviation of data points. Local fairing methods such as cardinal splines

method[3], knot insertion and deletion method[4] are to select 'bad points' and modify them. To be specific, some "bad points" such as redundant inflection point, discontinuity point of curvature etc. should be selected according to certain rules first, and then the "bad points" are to be modified to achieve smoothing results. But these local methods often require manual intervention, because "bad points" sometimes need to be determined by designers. The pretreatment for set of data points has been studied in the paper, so the whole fairing method has been applied in order to achieve automated computer data point smoothing, reducing manual intervention.

Whole fairing method is a complex nonlinear optimization problem. Meanwhile, intelligent optimization methods such as genetic algorithm have been widely used for solving nonlinear optimization problems. Genetic algorithm is easy to use in most cases, and it has strong robustness in solving nonlinear problem. Genetic algorithm also has other advantages which are in [3], random search can be carried out to obtain optimal results through simulating the process of biological evolution. Genetic algorithm has been widely used in the design of curves and surfaces [6-10]. Yoshimoto [5-6] has applied the genetic algorithm in curve fitting to planar data points and spatial data points, he has developed the key steps of genetic algorithm in fitting. A. Markus: [3] regarded the integral of curvature square as the objective function to obtain smooth interpolation B-spline curves through genetic algorithm. T. Hermann [9] and YH Chen [10] also applied the genetic algorithm to fair data points, but their method lacked numerical examples to verify their usefulness.

In this paper, on the basis of A. Markus's method [3] and the genetic algorithm, the discrete form of curve's energy is also used in the faring process. The discrete energy function of data points and the total deviated quadratic distance for control points before and after faring have been regarded as the optimization target. Researchers also develop the detection algorithm of feature points and the point set subdivision algorithm to control the size of population in per iteration. Finally, the algorithm will be tested and the test results will be discussed and analyzed.

II. FARING ALGORITHM BASED ON THE GENETIC ALGORITHM

Genetic Algorithm has made rapid development in the past 30 years, and experimental experience has been accumulated in the years of applications. For the size of the population, when the group is too small, it is clear that there will be inbreeding in population, and will produce morbid genes. When the population size is too large, genetic algorithm is difficult to convergence to best result, and robustness of genetic algorithm will decline. Large size of population often results a lot of useless operation, so it is a huge waste. In general sense [5-6,8-9], the population size should be controlled in less than 100 chromosomes to produce better optimization structure.

In this paper, the set of data points will be divided into smaller sets to ensure that each set of data points has two following characteristics:

1. There are as few feature points as possible in a faring point set. Inflection point and curvature extreme points are regarded as feature points of the point set. As for the data point sets, the fewer feature points mean that it is a faster processing speed in the fairing.

2. The number of data points in a set for faring should be less than 100.

A. Algorithm of finding feature points

First, researchers use circular interpolation for data points to compute the discrete curvature of each point. I.e. researchers interpolate three adjacent data points using an arc. It is shown in the below Figure :
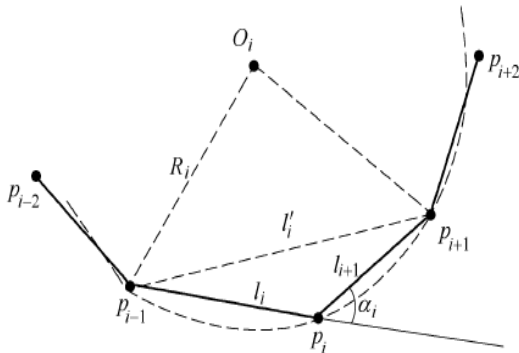

Figure 1.   Arc for interpolation for three adjacent data

The discrete curvature κ_i of each point is the curvature of this point on the interpolation circle. Then researchers can find all the feature points according to the discrete curvature of each point. If a point with its two adjacent points meets one of the following conditions:

(1) $\kappa_i \kappa_{i-1} < 0$ and $\kappa_i \kappa_{i+1} < 0$.

(2) $\Delta\kappa_i \Delta\kappa_{i-1} < 0$ and $\Delta\kappa_i \Delta\kappa_{i+1} < 0$ .Where $\Delta\kappa_i = \kappa_i - \kappa_{i-1}$, $\Delta\kappa_i = \kappa_{i+1} - \kappa_i$

Then this point is regarded as the feature point.

B. Subdivision algorithm of point set

Input: The original set of data points $\{T\}$

Output: Smaller sets of data points $\{T_1, \cdots, T_m\}$

Step 1: Find all the feature points in original data point set, denote them as $P = \{\bar{P}_1, \cdots, \bar{P}_l\}$.

Step 2: For $k = 1 : m$

Calculate the number M_k of data points between two adjacent feature points $\bar{P}_{k-1}, \bar{P}_k$;

If $M_k \leq 100$, turn to step 4; else If $M_k > 100$, turn to step 3.

Step 3: Insert a new feature point in these two the feature points and update feature point set P . The set T_i is separated to two sets.

Step 4: Output all feature points of the data point set. T_i is the point set contains all of the original data points between two adjacent feature points $\bar{P}_{i-1}, \bar{P}_i$.

C. Fairing algorithm for a single set of data points

Input: Point set with m data points $\{Q_i\}$ $i = 0, L, m$

Output: Point set of data point $\{P_i\}$ $i = 0, L, m$ after faring

Step 1: Initialize the population. Generate N chromosomes $\{G_i\}$ randomly which consists of the initial population.

Step 2: Compute the individual's fitness to evaluate each individual.

Step 3: Selection. Individuals with higher fitness in the current population will be inherited to the next group by a certain rule.

Step 4: Crossover. To produce a new individual by the probability of Pc .

Step 5: Mutation. To produce a new individual by the probability of Pv .

Step 6: Repeat Step 2-Step 5, output fair set of data points after meeting the terminating conditions.

III. DETAILS OF THE GENETIC ALGORITHM

In this section, researchers will describe all the details of genetic algorithm in fairing data point set.

A. Initial population

As for genetic algorithm, the initial population can be generated randomly. In order to match data points with chromosomes, the paper uses a binary code. And because that researchers need to control the size of the population first, so the size of initial population is fixed. Denote the initial population as $G = \{G_1 L \ G_K\}, K \leq 100$

B. Fitness function

Given m data points $\{Q_i\}$ $i = 0, L, m$, if there is a smooth curve passing through these data points, researchers can determine these data points are smooth. The following criteria can be used to measure the smooth degree of curve:

1) The second-order geometric continuity (Continuous position, continuous tangent direction and continuous curvature vector, which is denoted as G)

2) There is no singular point and inflection point

3) Even change in curvature.

4) The strain energy is smaller.

Researchers apply these criteria to develop local fairing method and whole fairing method. As for local fairing method, researchers should find out the redundant singular points, redundant inflection points and curvature jump points, and then researchers make an adjustment on these points to ensure the smoothing curve passing through them in the local area. Through adjusting the data point onto a smooth curve segment researchers achieve smooth data points. The energy method is mainly used in whole faring. In general, strain energy of a curve $C(u)$ can be described as the following form

$$E = \int_0^1 \kappa^2 ds \qquad (1)$$

$$\kappa(u) = \frac{\dot{C}(u) \times \ddot{C}(u)}{|C(u)|^3} \qquad (2)$$

Strain energy of curve can be calculated using integral. If researchers use this continuous model, it will increase the difficulty of computing. Therefore, researchers will adopt discrete s model, using

$$\sum_{i=1}^{n-1} \frac{1}{l_i + l_{i+1}} \|q_{i+1} - q_i\|^2 \qquad (3)$$

to substitute (1) Where $q_i = \dfrac{Q_i - Q_{i-1}}{l_i}$, $l_i = \|Q_i - Q_{i-1}\|$.

In pursuit of the curve while minimizing the strain energy, researchers should also make the deviation between the data point after smoothing adjustment and the original data points become small as much as possible, so researchers add the distance min-term

$$\sum_{i=1}^{n} \beta_i \|P_i - Q_i\|^2 \qquad (4)$$

The optimization goal is

$$E = \alpha \sum_{i=1}^{n-1} \frac{1}{l_i + l_{i+1}} \|q_{i+1} - q_i\|^2 + \sum_{i=1}^{n} \beta_i \|P_i - Q_i\|^2 .$$

C. Selection

In this algorithm, roulette wheel selection is applied in selection. If the fitness value of individual fitness G_i is f_i, the probability of this individual being selected is computed by $p_i = \dfrac{f_i}{\sum\limits_{i=1}^{n} f_i}$.

D. Crossover

Researchers use the linear combination crossover in this process. Chromosome has probability Pc been selected to choose for crossover.

Set up two fathers chromosome G_1 and G_2, r is a random number generated in $[0,1]$. The next generation generated through crossover is as follows:

$$G_1' = rG_1 + (1-r)G_2$$
$$G_2' = (1-r)G_1 + rG_2$$

E. Mutation

In this paper, an individual is selected by the probability of Pv for variation. Variation is an indispensable part of the genetic algorithm; the specific process is as follows:

A random number b is generated uniformly in $[0,1]$, and if $b < Pv$, variation happens.
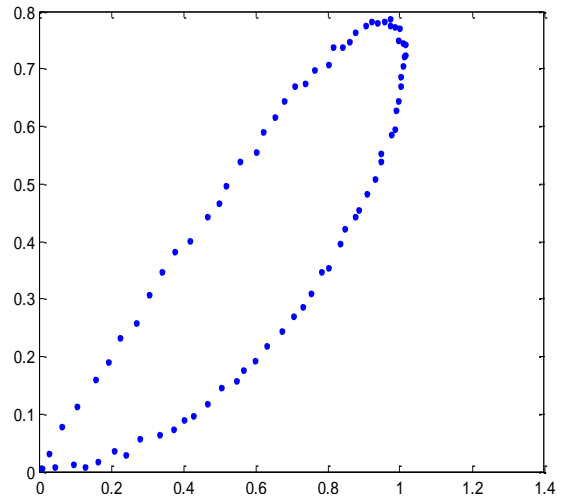
IV. EXAMPLE

Figure 2. Scanned data points with errors
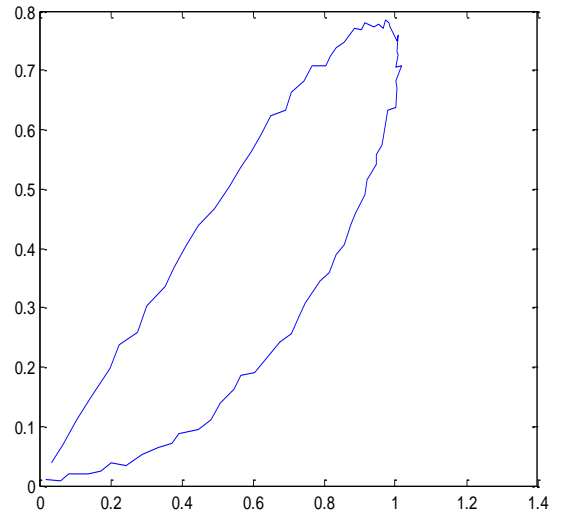

Figure 3. Reconstruction curve from error data points
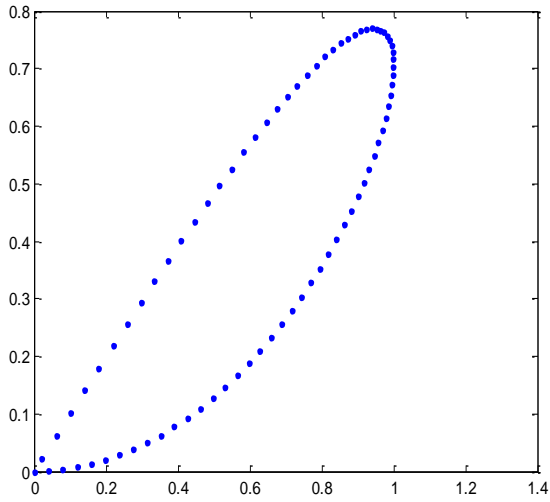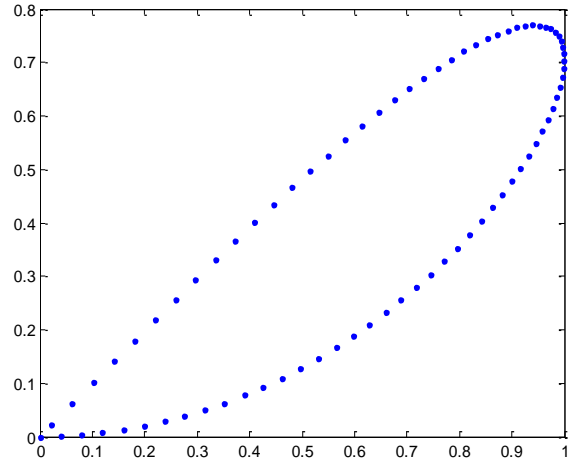
Figure 4. Data points after fairing

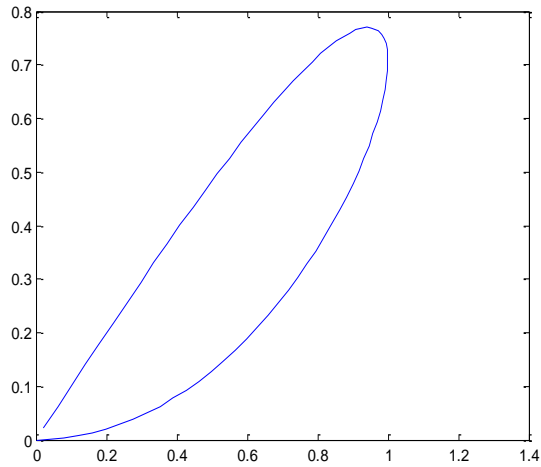
Figure 5. Smooth curve reconstrution from fair data points

Example1. In this numerical example, Fig. 2 shows a scanned set of data points with noise, this set of data point contains 93 data points. Fig. 3 shows a curve reconstructed from the original data points with noise. Data points after fairing are shown Fig. 4. The final construction curve is shown Fig. 5, and researchers can see that it is a smooth curve passing through these data points.

Figure 6. Data point set with larger scale

Figure 7. First part of fairing data points

Researchers take crossover probability $Pc = 0.8$ and mutation probability $Pv = 0.05$, and researchers get the smoothed data points after the 13th iteration. Fig. 5 shows the smooth curve of this algorithm.
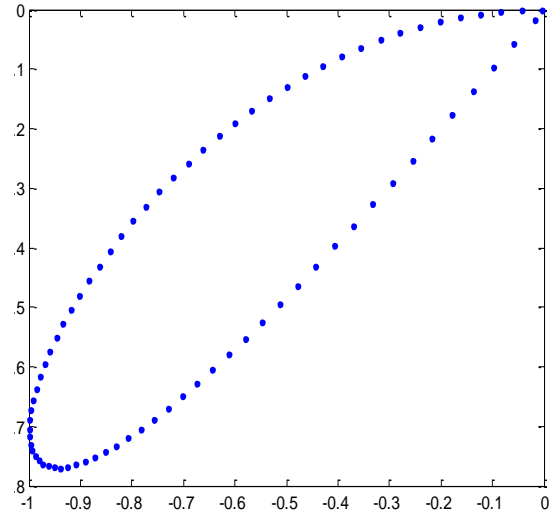

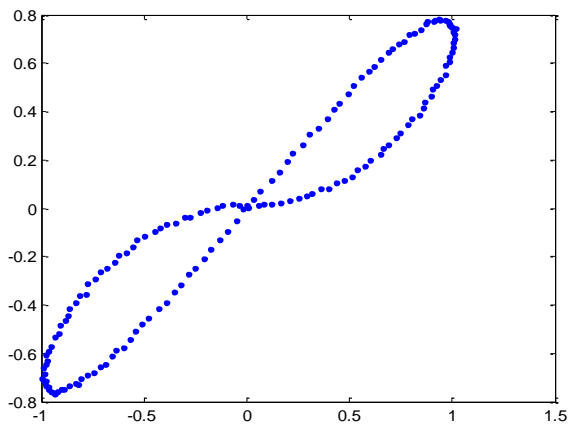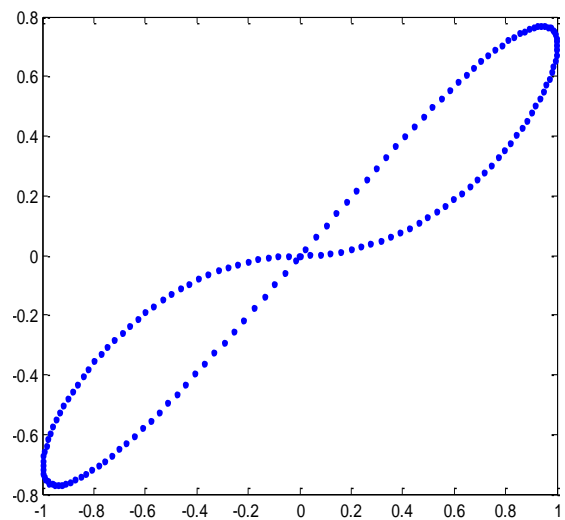Figure 8. Second part of fairing data points

Figure 9. Whole fair data points

Example 2. In this example, this data point set contains 193 data points, so the scale of set of data points is larger than 100. If researchers use the same crossover and mutation probability, genetic algorithm needs to iterate 38 times, and it outputs data points in the fair set with the same bound of error. According to the algorithm put forward in the paper, the data point set T is divided into two sets T_1, T_2, containing 100 and 93 data points respectively. If other conditions are same, researchers only need 12 iterations to get the final fair data points, so it is less than half of times that for faring 193 data points once a time. Genetic algorithm has been applied for fairing two sets of data points, and fair data points are shown in Fig. 7 and Fig. 8. The final fair set shown in Fig. 9 consists of two smaller sets in Fig. 7 and Fig. 8.

V. CONCLUSIONS

In this paper, the genetic algorithm is applied to fair data point set, and corresponding set subdivision algorithm is also designed for different sizes of sets. If the size of data points is large, data point set should be divided into several smaller sets. Then researchers fair each point set by the genetic algorithm respectively, and the final fair point set contain all smaller fair point sets. In this paper, the key steps of genetic algorithm are also described and a new fitness function is designed to fair data points. Experimental results show that the effect of the smoothing of the algorithm is better. However, the genetic algorithm has the disadvantage of converging to local optimal value and time-consuming, so how to design appropriate steps to reduce the time of a single iteration is the direction of further research. As computer performance improves quickly, the algorithm will play an important role in the faring for set of data points.

REFERENCES

[1] C.Y Deng, H.G Lin, Progressive and iterative approximation for least squares B-spline curve and surface fitting, Computer-Aided Design, Volume 47, February 2014, pp 32-44.

[2] L.A.Piegl, W.Tiller. Parametrization for surface fitting in reverse engineering, Computer-Aided Design,2001,33(8):pp593-603

[3] A.M Bica. Optimizing at the end-points the Akima's interpolation method of smooth curve fitting Computer Aided Geometric Design, Volume 31, Issue 5, June 2014, Pages 245-257.

[4] X.N Yang*, G.Z Wang,Planar point set fairing and fitting by arc splines. Computer-Aided Design 33 (2001) 35-43.

[5] F. Yoshimoto,M.Moriyama. Data finting with a spline function automatic knot placement by a genetic algorithm. Trans. On Information Processing Society of Japan,1998,39(9): 2572-2580.

[6] F. Yoshimoto,T. Harada,Y. Yoshimoto. Data finting with a spline using a real-coded genetic algorithm. Computer Aided Design . 2003,35(8): 751-760.

[7] G Renner, A Ekart. Genetic algorithms in CAD, Computer Aided Design . 2003,35(8):707-708.

[8] J. Lampinen. Cam shape optimization by genetic algorithm. Computer Aided Design . 2003,35(8):727-737.

[9] T. Hermann, J. Peters, T. Strotman.A geometric constraint on curve networks suitable for smooth interpolation. Computer-Aided Design, Volume 43, Issue 7, July 2011, Pages 741-746

[10] Y. H. Chen,C. Y. Liu.Quadric surface extraction using genetic algorithms. Computer Aided Design . 1999,31(1):101-107.