Arabic Fonts Representation in Cubic Bézier Curve using Different Soft Computing Algorithm

N F I C Amat*, Z R Yahya, N A Rusdi, N A Helmee and W Z A W Muhamad

*Corresponding author: fatiahizzati@gmail.com

Institute of Engineering Mathematics, University Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia

Abstract. This study presents the comparison between two meta-heuristic algorithms for curve reconstruction of Arabic fonts. The two algorithms namely Sine Cosine Algorithm (SCA) and Whale Optimization Algorithm (WOA) are used for comparison purposes. These algorithms have been applied to optimize the solution of curve fitting problem by using cubic Bézier curve. There are several pre-processing steps involve before curve fitting process has been done. Firstly, extract the boundary. Then, detect the corner of outline fonts and the third step is chord length parameterization. The aim of this study is to find the best curve that fit the data. Sum Square Errors (SSE) has been used to calculate the error given by original image and parametric curve. The finding of this study shows that, WOA gives better curve approximation based on the errors result than SCA. This algorithm significantly improves the shape of the corner of the outline Arabic fonts.

1. Introduction

Curve reconstruction is a crucial field in computer graphic especially for the curves which are not easily computed. Generally, it is a process of reconstructing a curve or a mathematical function, which has the best fit to a set of data points [1]. The main challenging parts of curve reconstruction are to get the smoothness and accurateness of the curve [2]. It has been grabbed much attention by researchers in the last few years [3–6]. There are numerous aspects of curve reconstruction such as font designing[7–10], data visualization[11], capturing hand-drawn images on computer screens [6] and medical image [12]. The important phase of curve reconstruction is curve fitting. The goal of curve fitting is to spot the set of data points that can approximately represent the given curve. Before curve fitting has been done, a few pre-processing phase (boundary extraction, corner point detection and chord length parameterization) needed in order to interpolate the data.

There are some frequently curves had been chosen by the researchers such as B Spline[3], cubic Spline[13, 14], cubic Bézier curve [4–6, 12, 15], and quartic Bézier curve [9] for solving curve fitting problems. Cubic Bézier curve had been chosen as an appropriate curve because of its advantages and interesting properties which make it differ from other curves [13]. The control points of the curve can be adjusted to effectively produce a good result.

Soft Computing is basically an optimization technique to find the best solution in solving various types of problem especially in complex engineering problems [1]. There are several example of soft computing technique such as Genetic Algorithm (GA) [2, 16], Differential Evolution (DE) [1, 10, 17], Artificial Bee Colony algorithm (ABC) [5, 12, 18], Whale Optimization Algorithm (WOA) [6, 19] and Sine Cosine Algorithm (SCA) [20]. According to [5], Artificial Bee Colony Algorithm has been used to reconstruct the Arabic font and presents great result since the fitted cubic Bézier curve approximates the original image. Based on [6], WOA proposed to explore curve fitting for handwriting image and the error obtained quite large but most of the fitted cubic Bézier curve is on its original boundary. SCA is a simple effective optimization algorithm based on sine and cosine mathematical functions developed by Mirjalili to solve optimization problems. Numbers of researches shown SCA has been successfully employed in many optimization problems such as feature selection [21], manuscript image binarization [22], modeling optimization [23] and data clustering [24]. In this study, the
comparison of two different soft computing algorithms which are SCA and WOA with cubic Bézier curve has been discovered in order to find the best curve in solving curve fitting problem.

This paper can be divided into six sections. Section 2 discussed the Bézier curve. Then it was followed by explaining the process of SCA and WOA. Next, in section four the pre-processing process will be discussed in details. The elaboration on curve fitting process for both algorithms had been done in Section 5. Lastly, the paper has been concluded in Conclusion section.

2. Cubic Bézier Curve
Bézier curve was introduced by the French engineer Pierre Bézier are widely used in computer graphics to model smooth curves [25]. Using the Bernstein polynomials as a basis function, a Bézier curve can be defined of degree $n$ (in terms of Bernstein polynomials) to be a curve in parametric form $S(t)$. The general form of Bézier curve of degree $n$ is represented by

$$S(t) = \sum_{i=0}^{n} B^n_i(t) P_i \quad 0 \leq t \leq 1 \quad (1)$$

where $P_i$ is a control point and $B^n_i$ are blending function, known as Bernstein polynomial basis function and $t$ is parameter. Basis function of Bernstein is then defined as

$$B^n_i(t) = C(n-i), 0 \leq t \leq 1 \quad (2)$$

In this study, Bezier curve of degree three was used. Notes that, cubic Bézier curve is determined by four control points $P_0, P_1, P_2$ and $P_3$. From (1), cubic Bézier curve can be illustrated as $S(t) = \sum_{i=0}^{3} B^3_i(t) P_i$.

3. Sine Cosine Algorithm (SCA)
This chapter discussed the Sine Cosine Algorithm (SCA), which is a recently meta-heuristic that applies a model based on trigonometric functions to estimate the global optima of optimization problems [20]. The optimization process is generally split into two phases: exploration and exploitation. In each iteration, the position update of $X^i_t$ formula can be obtained by the following equation:

$$X^{i+1}_t = \begin{cases} 
X^i_t + r_1 \sin(r_2) | r_3 P^i_t - X^i_t | & r_4 < 0.5 \\
X^i_t + r_1 \cos(r_2) | r_3 P^i_t - X^i_t | & r_4 \geq 0.5 
\end{cases} \quad (3)$$

where $X^i_t$ represents the position of current solution in $i$ th dimension at $t$ th iteration, $P^i_t$ is position of the destination point in the $i$ th dimension at the $t$ th iteration, while $\| \cdot \|$ specifies the absolute cost. SCA used 4 main random parameters namely $r_1, r_2, r_3$ and $r_4$. These parameters which effectively balances the exploration and exploitation ability of the algorithm in solution space. Parameter $r_1$ decides that search agent is going to do whether exploration or exploitation. Then, parameter $r_2$ decides how far the solution’s movement should be towards or outwards the destination. In order to stochastically emphasize ($r_3 > 1$) or deemphasize ($r_3 < 1$) the effect of destination in finding the distance, parameter $r_3$ assign a random weight for the destination. Finally, $r_4$ ranging from 0 to 1 decided whether sine or cosine formula is going to be used.
The parameter $r_1$ is the main stochastic variable defining the movement direction of the solution towards or outwards the destination. In population-based algorithms to converge a point at the end of optimization, the exploration phase should be decreased. The parameter linearly $r_1$ decreases in SCA as follows:

$$r_1 = a - \frac{i}{T}$$  \hspace{1cm} (4)

where $i$ is the current number of iterations, $T$ is the maximum number of iterations, $a$ is a constant.

4. Whale Optimization Algorithm (WOA)
Whale Optimization Algorithm (WOA) is a bio-inspired meta-heuristic algorithm inspired by hunting mechanism of humpback whales [19]. This special hunting mechanism called bubble net feeding method. This algorithm includes three typical behavior of humpback whale and models them mathematically. The mathematical model of encircling prey, spiral bubble-net feeding maneuver and search for prey is described in the following section:

4.1. Encircling Prey
Humpback whales at first will spot the position of prey and then they encircle the prey. Target prey or close to the optimum is assumed as current best candidate solution. The other search agents will hence try to update their position towards the best search agent after the best search agent is defined. This action is demonstrated by the following equations:

$$\bar{D} = |\bar{c} \cdot \bar{x}^i(t) - \bar{x}(t)|$$  \hspace{1cm} (5)
$$\bar{x}(t + 1) = \bar{x}^i(t) - \bar{A} \cdot \bar{D}$$  \hspace{1cm} (6)

where $t$ indicates the current iteration, $\bar{A}$ and $\bar{c}$ are coefficient vectors, $\bar{x}^i$ is the position vector of the best solution, $\bar{x}(t)$ is position vector, $\bar{D}$ is distance between $i^{th}$ whale and the prey, $| \cdot |$ is absolute cost, and $\lfloor \cdot \rfloor$ is an element wise multiplication. $\bar{x}^i$ should be updated in each iteration if there is a better solution. The vectors $\bar{A}$ and $\bar{c}$ are calculated as follows:

$$\bar{A} = 2\bar{a} \cdot \bar{r} - \bar{a}$$  \hspace{1cm} (7)
$$\bar{c} = 2 \cdot \bar{r}$$  \hspace{1cm} (8)

where $\bar{a}$ is linearly decreased from 2 to 0 over the course of iterations (in both exploration and exploitation phases) and $\bar{r}$ is a random vector in $[0,1]$.

4.2. Bubble-net Attacking Method (Exploitation Phase)
There are two approaches in this hunting method that has been mathematically model.
1. Shrinking encircling mechanism: This technique employed by decreasing linearly the value of $\bar{a}$ from 2 to 0. Random value for a vector $\bar{A}$ in the range between $[-1,1]$.
2. Spiral updating position: Mathematical spiral equation for position update between humpback whale and prey that was helix-shaped movement as follows:

$$\bar{x}(t + 1) = \bar{D}' \cdot e^{bt} \cdot \cos(2\pi l) + \bar{x}^i(t)$$  \hspace{1cm} (9)
where \( l \) is a random number \([-1,1]\), \( b \) is constant defines the logarithmic shape, \( \vec{D} = \vec{X}'(t) - \vec{x}(t) \) express the distance between \( i^{th} \) whale to the prey (best solution so far).

Note: Assume that there is a probability of 50% that whale either follow the shrinking encircling or logarithmic path during optimization. The mathematical model as follows:

\[
\vec{x}(t + 1) = \begin{cases} 
\vec{X}'(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\
\frac{\vec{D}'}{e^{bl} \cdot \cos(2\pi l)} + \vec{X}'(t) & \text{if } p \geq 0.5
\end{cases}
\]  

(10)

where \( p \) is a random number in \([0,1]\).

4.3. Search for Prey (Exploration Phase)

The vector \( \vec{A} \) can be used for exploration to search for prey. Vector \( \vec{A} \) also used random values greater than 1 or less than -1. At this phase, the update position of search agent is according to a randomly chosen search agent. If \( |\vec{A}| > 1 \) emphasize exploration where a random search agent is chosen and it allow the WOA algorithm to perform a global search. If \( |\vec{A}| < 1 \) emphasize best solution is selected for updating the position of the search agent. The mathematical model is as follows:

\[
\vec{D}' = |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{x}|
\]

(11)

\[
\vec{x}(t + 1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D}
\]

(12)

where \( \vec{X}_{\text{rand}} \) is a random position vector (a random whale) chosen from current population.

5. Pre-Processing Steps

There are several pre-processing phases involved before the curve fitting process, which include boundary extraction, corner point detection and chord length parameterization. This section will explain in detail the role of each phase. The pre-processing phases and curve fitting process can be summarized in a flowchart as shown in Figure 1.

**Figure 1.** Flow chart pre-processing phases and curve fitting process

5.1. Boundary Extraction and Corner Point Detection

The first process to reconstruct the curve starts with the boundary extraction of the original image. The boundary of the image is extracted to represent an object’s shape in graphical representation (number of boundary points) [17]. There are numerous algorithms has been used to extract the boundary of the generic shape such as [26, 27]. In this work, boundary of digitized image is extracted by using MATLAB function that called Boundaries in order to extract the boundary of the font images. The
function returns number of pieces and for each piece number of boundary points and values of these boundary points \( P_i = (x_i, y_i), i = 1, ..., n \).

The corner points of the image are detected, after finding out the boundary points. Corner points are defined as the point that separates the outline into various segments before curve fitting. Corner points play an important role in image representation [7] and preprocessing phase of outline capturing systems [29]. Various corner detection algorithms have been discovered such as Davis Algorithm, Harris Corner Detection Algorithm, Rosenfeld and Johnston (RJ73) Algorithm and SAM06 Algorithm [30]. In this paper, SAM06 algorithm was applied to detect the corner points. The boundary extraction and corner detection of the Arabic font are represented in Table 1.

**Table 1.** Boundary extracted and detection of corner points.

| Name  | Lam Alif | Kha | Ka |
|-------|----------|-----|----|
| Original Images | ![Lam Alif](image) | ![Kha](image) | ![Ka](image) |
| Boundary Extraction and Corner Detections | ![Lam Alif](image) | ![Kha](image) | ![Ka](image) |

5.2. **Chord Length Parameterization**

In section 2, cubic Bézier curve is defined by the beginning point and the ending points, which are interpolated and have two interior points that control the shape of the curve. This cubic Bezier curve consists of four control points: \( P_0, P_1, P_2 \) and \( P_3 \). The initial and final points of the curve are \( P_0 \) and \( P_3 \) respectively. The interior control points \( P_1 \) and \( P_2 \) are used to control the shape of the curve. The cubic Bezier curve can be written in a matrix form by expanding the analytic definition of the curve into its Bernstein polynomial coefficients, and then writing these coefficients in a matrix form using the polynomial power basis [25]. That is,
Chord-length parameterization which been receiving attention in CAGD is commonly adopted for simulating curve arc-length parameterization in discrete point data interpolation and approximation [5]. Chord-length parameterization is used to estimate the parametric value \( t \) associated with each point \( p_i \) as follows:

\[
S(t) = \sum_{i=0}^{3} B_i(t)P_i
\]

\[
S(t) = B_0^3(t)P_0 + B_1^3(t)P_1 + B_2^3(t)P_2 + B_3^3(t)P_3
\]

\[
= (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3
\]

\[
= [(1-t)^3 3t(1-t)^2 3t^2(1-t) t^3 ] \begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix}
\]

Chord-length parameterization which been receiving attention in CAGD is commonly adopted for simulating curve arc-length parameterization in discrete point data interpolation and approximation [5]. Chord-length parameterization is used to estimate the parametric value \( t \) associated with each point \( p_i \) as follows:

\[
t_i = \begin{cases} 
0 & \text{if } i = 1 \\
\frac{p_1p_2 + p_2p_3 + \ldots + p_{i-1}p_i}{p_i p_2 + p_2p_3 + \ldots + p_{n-1}p_n} & \text{if } i \leq n-1 \\
1 & \text{if } i = n
\end{cases}
\]  

(14)

The goal is to approximate the boundary of the original image by a parametric curve in an optimal way.

6. Fitting the Curve using Sine Cosine Algorithm (SCA) and Whale Optimization Algorithm (WOA)

The main part of curve reconstruction is called curve fitting. The three selected Arabic fonts Lam Alif (ٌ), Kha (خ) and Kaf (ک) were employed in this study. The aim of this research is to compare the results obtained by using SCA and WOA in curve fitting phase. The same pre-processing steps were adopted in order to interpolate the data. At this phase, both soft computing algorithms were applied to approximate the two middle points \( P_1 \) and \( P_2 \). Then, these points will be used to find the parametric curve that gives the best optimal results for given values of \( t \).

There are some parameters included in this phases which are population size, number of iterations, limit, and dimension. In this study, population size is set to be 50 with 100 numbers of iterations for both algorithms. The middle points \( P_1 \) and \( P_2 \) consist of \((x_1, y_1)\) and \((x_2, y_2)\), respectively. So, the dimension is set to be four since it expressed \( x_1, x_2, y_1 \) and \( y_2 \).

The fitted cubic Bézier curve is obtained, when the best value of points \( P_1 \) and \( P_2 \) are found. The goal of this study is to minimize the distance between boundaries of the original images and fitted cubic Bézier curve obtained by using SCA and WOA. In order to achieve the different result, Sum Square
Error (SSE) has been used to calculate errors between the parametric points of generated curve and original image. The result of SCA will be compared with WOA. The final results presented visually and numerically are illustrated in Figure 2 and Table 2 respectively. Figure 2 shows the demonstration of fitted cubic curve (blue line) over the original curve (dotted line). The fit of the cubic Bézier curve is on boundary of bitmap image. The reconstruction have similar image visually while the Table 2 represents the errors obtained between Sine Cosine Algorithm (SCA) and Whale Optimization Algorithm (WOA). The result shows that the performance of WOA is better than SCA since it produced a small error.

**Figure 2.** The visualization result of fitted cubic Bézier curve (blue line) over the original curve (dotted line)

**Table 2.** The error obtained by each image.

| Image   | Sine Cosine Algorithm (SCA) | Whale Optimization Algorithm(WOA) |
|---------|-----------------------------|-----------------------------------|
| Lam Alif (Ʌ) | 10.3980                    | 10.1805                           |
| Kha (خ)      | 5.5936                     | 2.6588                            |
| Ka (گ)       | 16.8602                    | 16.6891                           |

7. Conclusion
This paper applied two different soft computing algorithms to solve curve fitting problem. The two algorithms are SCA and WOA. Sum Square Error (SSE) was used in order to estimate the errors between the boundary of bitmap images and the parametric curve. This study shows that WOA is more effective method than SCA in solving this problem since it produced an accurate representation with small error. The result illustrated, the reconstruction have similar image visually, while the error shown that the WOA lead all from the three selected image. In conclusion, both algorithms remain as the interesting algorithm that can be explored in more details and can be applied in numerous problems.

Acknowledgement
The authors highly acknowledge the financial sponsorship provided by Fundamental Research Grant Scheme FRGS/1/2017/STG06/UNIMAP/03/2.

References
[1] Hasan Z A, Yahya Z R, Rusdi N A and Roslan N 2018 Curve Reconstruction In Different Cubic Functions Using Differential Evolution *MATEC Web of Conference* **150** 06030
[2] Majeed A and Piah A R M 2014 Image reconstruction using rational ball interpolant and genetic algorithm *Applied Mathematical Sciences* **8**(74) 3683-3692
[3] Haron H, Rehman A, Adi D I S and Saba T 2012 Parameterization Method on B-Spline Curve *Mathematical Problems in Engineering* **640472**
[4] Rusdi N A and Yahya Z R 2015 Reconstruction of generic shape with cubic Bézier using least
square method AIP Conference Proceedings, AIP Publishing 1660 050004
[5] Rusdi N A and Yahya Z R 2016 Reconstruction of Arabic Font using Artificial Bee Colony Algorithm ARPN J Eng Appl Sci 2016 11 10761–10767
[6] Helme N A, Yahya Z R and Hasan Z A 2018 Implementation of whale optimization algorithm for handwriting representation AIP Conference Proceeding AIP Publishing 2013 020007
[7] Yahya F, Md Ali J, Majid A A and Ibrahim A 2006 An automatic generation of G1 curve fitting of arabic characters International Computer on Graphics, Imaging and Visualisation: Techniques and Applications, CGIV’06. IEEE, 542-547
[8] Poon S H and Majid B A 2016 Interpolation between Outline Font Characters 2016 First International Conference on Multimedia and Image Processing (ICMIP) IEEE 61–66
[9] Rusdi N A and Yahya Z R 2015 Reconstruction of Arabic Font with Quartic Bézier Curve (Pembinaan Semula Fon Arab menggunakan Lengkung Bézier Kuartik) Sains Malaysiana 44 1209–1216
[10] Roslan N and Yahya Z R 2015 Reconstruction of Font with Cubic Bezier Using Differential Evolution Sains Malaysiana 44 1203–1208
[11] Yahya Z R, Piah A R and Majid A 2012 Curve Fitting Using Conic by Evolutionary Computing Journal of Mathematics and Statistics 8 107-110
[12] Rusdi N A, Yahya Z R, Roslan N, Muhamad W and Azman W Z 2018 Reconstruction of Medical Images Using Artificial Bee Colony Algorithm Mathematical Problems in Engineering 8024762
[13] Sarfraz M and Al-Dabbous N A 2013 Curve representation for outlines of planar images using multilevel coordinate search WSEAS Transactions on Computers 12 62–73
[14] Irshad M, Sarfraz M and Hussain M Z 2015 Reverse Engineering of Planar Objects Using Imperialist Competitive Algorithm 19th International Conference on Information Visualisation. IEEE, 430–435
[15] Roslan N and Yahya Z R 2016 Parameterization method on cubic Bézier curve fitting using differential evolution AIP Conference Proceeding, AIP Publishing 1775 030075
[16] Sarfraz M, Irshad M and Hussain M Z 2013 Reverse Engineering of Planar Objects using Genetic Algorithm Sains Malaysiana 42 1167-1179
[17] Hasan Z A, Yahya Z R and Piah A R M 2015 Differential Evolution for Reverse Engineering Using Cubic Ball Curve Educ J Sci Math Technol 2015 2 58–68
[18] Rusdi N A and Yahya Z R 2016 Artificial Bee Colony algorithm for curve reconstruction AIP Conference Proceeding, AIP Publishing 1775 030077
[19] Mirjalili S and Lewis A 2016 The Whale Optimization Algorithm Advances in engineering software 95 51-67
[20] Mirjalili S 2016 SCA: A Sine Cosine Algorithm for solving optimization problems Knowledge-Based Systems 96 120–133
[21] Hafez A I, Zawbaa H M, Emary E and Hassanien A E 2016 Sine cosine optimization algorithm for feature selection International Symposium on Innovations in Intelligent Systems and Application INISTA, IEEE. 1–5
[22] Elfattah M A, Abuelenin S, Hassanien A E and Pan J S 2016 Handwritten Arabic manuscript image binarization using sine cosine optimization algorithm Advances in Intelligent Systems and Computing, Springer, Cham 273–280
[23] Kaur S and Prashar S A 2016 A novel sine cosine algorithm for the solution of unit commitment problem International Journal of Science, Engineering and Technology Research 5 12
[24] Kumar V and Kumar D 2017 Data Clustering Using Sine Cosine Algorithm In Handbook of research on machine learning innovations and trends, IGI Global 715–726
[25] Pastva T A 1998 Bezier Curve Fitting Naval Postgraduate School Monterey CA
[26] Itoh K and Ohno Y 1993 A curve fitting algorithm for character fonts. Electronic Publication 6
[27] Sarfraz M and Khan M A 2002 Automatic outline capture of Arabic fonts. Information Sciences 140 269–281
[28] Jähne B 2005 Digital Image Processing New York: Springer-Verlag, Chapter 1
[29] Sarfraz M, Masood A and Asim M R 2004 A Web Based System for Capturing Outlines of 2D Objects Proc. of International Conference on Information & Computer Science, Dhahran, Saudi Arabia

[30] Sarfraz M 2008 Corner Detection for Curve Segmentation In: Interactive Curve Modeling