Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 16 No. 2
jtec.utem.edu.my
DOI: https://doi.org/10.54554/jtec.2024.16.02.004

Feature Extraction Algorithm based Metaheuristic Optimization for Handwritten Character Recognition

Muhammad Arif Mohamad[1], Muhammad Aliif Ahmad[2], and Jamilah Mahmood[2]
[1]*Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Pahang, Malaysia*
[2]*Faculty of Computing, Universiti Teknologi Malaysia, Johor Bharu, Johor, Malaysia*
arifmohamad@umpsa.edu.my

Article Info

Article history:
Received Apr 15th, 2024
Revised June 12th, 2024
Accepted June 29th, 2024
Published June 30th, 2024

Index Terms:
Metaheuristic
Honey Badger Algorithm
Feature Extraction
Freeman Chain Code
Handwritten Character Recognition

Abstract

Interest in feature extraction for Handwritten Character Recognition (HCR) has been growing due to numerous algorithms aimed at improving classification accuracy. This study introduces a metaheuristic approach utilizing the Honey Badger Algorithm (HBA) for feature extraction in HCR. The Freeman Chain Code (FCC) is employed for data representation. One challenge with using FCC to represent characters is that extraction results vary depending on the starting points, affecting the chain code's route length. To address this issue, a metaheuristic approach using HBA is proposed to identify the shortest route length and minimize computational time for HCR. The performance metrics of the HB-FCC extraction algorithm are route length and computation time. Experiments on the algorithm use chain code representations from the Center of Excellence for Document Analysis and Recognition (CEDAR) dataset, containing 126 uppercase letter characters. According to the results, the proposed HB-FCC method achieves a route length of 1880.28 and requires only 1.07 seconds to process the entre set of character images.

I. INTRODUCTION

The accuracy of handwritten character recognition (HCR) is influenced by the sensitivity of the chosen characteristics. There are now numerous feature extraction techniques that exist in the literature [1]. Feature extraction involves generating information features from raw data to maximize sample variability between classes while minimizing it within a class. Therefore, when representing each class, it is essential to identify a set of qualities or characteristics that can differentiate it from other classes while remaining unaffected by variations in traits [2]. The review of feature extraction techniques for character recognition by [3] provides a good analysis of these methods.

Chain code is a method for processing and analyzing images based on boundary feature representation. Freeman [4] introduced the first chain coding method, known as Freeman Chain Coding (FCC). The approach used to traverse the image and its starting point significantly impacts the difficulty of chain-coding procedure. Even if the image remains the same, the starting point of image characters produce different FCC directions. Characters' starting points were chosen arbitrary to obtain the best answer. However, there has been limited active investigation into chain code constructions via a single continuous path with minimized chain codes lengths. A metaheuristic optimization approach was used to extract a continuous chain code that accurately reflects handwritten characters to solve this problem.

In summary, this study proposed a metaheuristic algorithm-based chain code feature extraction that utilizing the Honey Badger Algorithm (HBA). The HBA was used to shorten the FCC. The main challenges in using FCC to represent character images are the revisit walks, node branches, and commencement points, all of which affect the length of the chain code. HBA is used to create an FCC that accurately depicts the image character, thereby addressing these problems.

II. RELATED WORKS

Handwritten character recognition (HCR) is often described in three stages: preprocessing, feature extraction, and classification. The objective of the preprocessing phase is to produce a character image free from impurities, facilitating efficient and direct feature extraction. The second phase aims to eliminate image data redundancy. The third phase involves identifying or recognizing the characters in the images. Various methods for character recognition are currently in use.

In this study, boundary extraction-based representation methodologies for image processing and pattern recognition were based on the FCC [5]. Chain code representation provides a character image border, with codes indicating the path traced subsequent pixels. As FCC has expanded and advanced, many different chain coding techniques have emerged [6]. There are two directions for chain code: four-neighborhood and eight-neighborhood methods. This study uses the eight-neighborhood method for character extraction.

To construct the FCC accurately and optimize its by minimizing its length, a metaheuristic optimization approach namely Honey Badger Algorithm (HBA). Proposed by [7], HBA was inspired by the foraging behavior of honey badgers. Honey badgers seek food in two ways: by using their own sense of smell or by following the honeyguide bird, which is known for finding honey readily.

HBA is distinguished by its balance between exploration and exploitation, which improves its capacity to handle difficult optimizations problems involving multiple local solutions [8]. The exploration and exploitation stages of HBA depict the dynamic search behavior of honey badgers, which involves digging and honey-finding strategies. Because of its various advantages over other methods, HBA is a preferred solution for many optimization situations.

One of the primary benefits of HBA is its ability to balance exploration and exploitation throughout the search process, resulting in higher convergences rates and better performance compared to competing algorithms [9]. Furthermore, [10] stated that HBA requires few parameters during the initialization step, making it simple to implement and less prone to overfitting. Furthermore, HBA's versatility and expandability enables it to be easily adjusted for various optimizations difficulties.

III. METHODOLOGY

The proposed methodology for the HB-FCC feature extraction algorithm is shown in Figure 1.

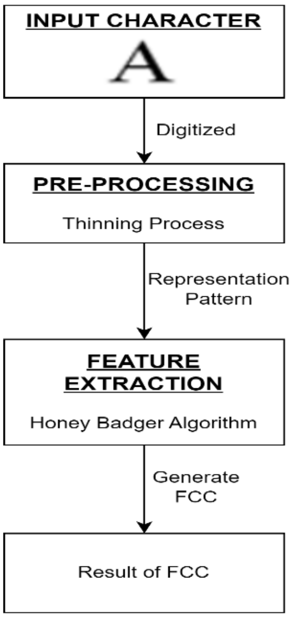
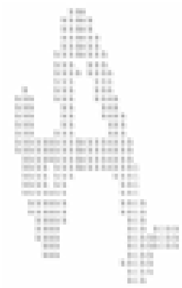


Figure 1. Proposed Methodology

The preprocessing phase involves only the thinning process. According to [11], character recognition was performed using features extraction with TBI. The CEDAR dataset is used as the data source for this experiment. This study focuses exclusively on the 126 upper-case Latin letters (A-Z). Table 1 displays samples of TBIs obtained through the thinning method.

Table 1
Sample of TBI Proposed by Engkamat

Raw Data	TBI Proposed by Engkamat
	

A. *Proposed HB-FCC Feature Extraction Algorithm*

HBA is used to extract chain code features based on thinned binary images (TBI) as illustrated in Table 1. The creation of chain codes based on binary images can be simulated using a graph problem technique. First, the binary images are converted into a digraph containing a vertex and an edge. The vertices of the network are defined based on the total number of single neighbor nodes and all nodes with more than two neighbors. A node with two neighbors and a connection to the starting vertex forms the network's edge. The lengths are determined by counting all the nodes that exist between any two vertices. Figure 2 shows the entire graph.

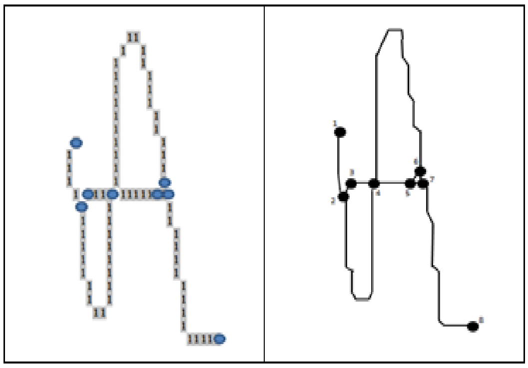


Figure 2. Character Transformation of Table 1 into Graph

The proposed HB-FCC extraction algorithm represents chain code solution in collection of edges. An edge is a representation of the solution, starting and ending at similar nodes. Two distinct edges may originate from the same node, or one edge may end at another of the same nodes. Due to the frequent need to revisit previously visited nodes, an edge may be visited twice, allowing the solution representations to encompass the entire tour. The total number of nodes that the chain code needs to visit (including revisits), starting from the initial nodes, is defined as the objective function. Figure 3 shows the proposed HB-FCC feature extractions procedure.

The Pseudocode of HB-FCC Extraction Algorithm

Input data and settings parameter values
Clean image
Enumerate junction/end as nodes and interconnecting routes.
Generate random node sequences.
Start HBA search with objective function()=path length
Initialize the population of Honey Badger with random position
Define intensity (I)
Evaluate the objective function of each honey badger position.
Save the best position
Calculate the intensity (I)
Update position
Evaluate new position.
Until stopping criterion is archive

Figure 3. Proposed HB-FCC Feature Extraction Algorithm

The algorithm works as follows:

i. Clean and break the image into paths.
ii. Enumerate the junctions/ends as nodes and interconnecting routes. At this point, generate chain code for the corresponding routes.
iii. Find vertex by obtaining list of junctions by testing neighbor population (i=2).
iv. Connect the junctions by testing each connected edge, recording the edge and junction at the other end.
v. Generate random node sequences for HBA population initialization.
vi. Start HBA with objective function being the path length. The goal is to find the shortest route possible. Path length is the sequence of routes to reach all nodes.

IV. RESULT AND DISCUSSION

This section discusses the outcomes and conclusions of the study. HBA is used to create the continuous FCC, a feature of the handwritten character image. The proposed HB-FCC feature extraction consists of ten replications, each with one hundred chain code solutions for each TBI. The results of the route length calculations obtained by HB-FCC for each character are shown as Tables 2.

Table 2
Generated route length by HB-FCC

CHAR	ROUTE LENGTH										AVG
	1	2	3	4	5	6	7	8	9	10	
A	136.40	150.60	134.80	137.60	137.60	133.80	130.40	141.80	130.80	126.60	136.04
B	103.20	105.00	97.80	100.20	102.20	99.40	98.60	100.20	93.80	98.20	99.86
C	63.00	59.80	61.20	62.00	60.20	62.80	63.40	62.60	66.40	60.00	62.14
D	67.20	67.20	67.20	67.20	67.20	167.20	67.20	67.20	67.20	67.20	67.20
E	81.20	84.40	85.40	84.80	86.80	78.20	84.80	83.80	91.80	86.20	84.74
F	64.60	63.20	61.80	66.20	61.80	63.20	64.80	67.20	63.00	64.00	63.98
G	74.60	73.80	83.60	72.40	76.40	79.80	73.40	75.00	79.00	68.40	75.64
H	123.40	128.00	106.40	126.80	118.20	118.00	115.40	129.40	118.00	116.80	120.04
I	55.75	53.50	56.50	58.75	55.25	59.25	56.00	53.00	59.75	57.25	56.50
J	60.80	63.00	61.60	63.60	61.40	63.00	61.60	61.20	61.20	62.60	62.00
K	90.00	86.60	83.40	91.00	100.40	92.80	95.00	92.20	83.20	94.00	90.86
L	45.00	42.80	43.60	46.00	40.80	40.80	41.40	41.60	41.20	40.40	42.36
M	84.00	79.40	74.00	80.20	80.00	80.60	81.40	78.00	86.20	76.80	80.06
N	49.60	49.20	49.00	50.00	48.80	49.20	48.40	49.20	48.80	49.40	49.16
O	64.20	64.00	64.20	64.00	64.00	64.00	64.00	64.00	64.00	64.00	64.04
P	112.50	111.50	112.25	117.00	112.50	112.75	101.50	113.00	106.75	124.75	112.45
Q	113.33	114.33	107.33	99.67	115.67	114.00	108.00	116.00	109.33	110.00	110.77
R	74.60	78.60	75.00	77.80	75.00	74.80	75.20	78.00	77.00	76.20	76.22
S	51.20	51.20	51.20	51.20	51.20	51.20	51.20	51.20	51.20	51.20	51.20
T	51.60	51.60	51.60	51.60	51.60	51.60	51.60	51.60	51.60	51.60	51.60
U	59.80	59.80	59.80	59.80	59.80	59.80	59.80	59.80	59.80	59.80	59.80
V	48.60	48.60	48.60	48.60	48.60	48.60	48.60	48.60	48.60	48.60	48.60
W	74.20	74.20	77.60	73.20	71.20	81.60	69.40	74.40	79.80	70.20	74.58
X	82.60	71.40	67.80	83.20	71.20	83.60	79.80	69.20	71.80	70.40	75.10
Y	46.40	44.60	44.40	44.80	45.20	44.80	45.60	45.80	45.20	45.80	45.26
Z	54.00	57.80	54.20	53.40	55.60	57.80	56.80	55.40	56.60	55.20	55.68

Table 3 presents the best, average, and worst outcomes of route lengths and computation times for all characters (A-Z). Based on the findings, the proposed HB-FCC gained route lengths of 1880.28 and it took just 1.10 seconds to process each binary image.

Table 3
Overall Result

	Route Length	Computation Time (s)
BEST AVERAGE	1880.28	1.07
WORST	1915.88	1.10
	1934.13	1.16

V. CONCLUSION

In this study, a metaheuristic technique called HB-FCC extraction, which uses the Honey Badger Algorithm (HBA) to extract features for handwritten character identification was proposed. The proposed HB-FCC extraction technique establishes a single continuous route and shortens the FCC lengths. The chain code was extracted using this method based on the TBI of handwritten characters. The recommended methodology includes representing the characters and applying the following solution representation methods: (1) character to graph transformations, (2) graph as a representation of the solution, and (3) the metaheuristic technique (HBA) to shorten the FCC.

The objective function of the solution representation is minimized by the proposed HBA-FCC algorithm. This objective function, which measures the effectiveness of an FCC solution, is defined by the number of nodes the FCC must visit from the initial node until all nodes are visited. HBA, then tries to find efficient FCC alternatives by leveraging its unique attributes to shorten the FCC. The calculation time and route length were chosen for this experiment since they depend on the starting point and are automatically modified by those factors as well as the number of times the chain code must be solved. The suggested HB-FCC extraction algorithm can extract and recognize such complex characteristics with relatively low computing effort and route length.

In conclusion, the number of recognizable image characters demonstrates the effectiveness of the FCC in representation. The route length created by HB-FCC for each character is shown in Table 2. Table 3 presents the results for all characters (A-Z) in terms of route length and computation time, showing the best, average, and worst outcomes.

For future work, the HBA can be further enhanced by hybridizing it with other metaheuristic algorithms, such as Biogeography-based Optimization (BBO) and Particle Swarm Optimization (PSO). The advantageous characteristics of HBA make it highly versatile for combining with other metaheuristic algorithms to produce hybrid metaheuristics and to apply them in various applications.

REFERENCES

[1] J. Paul, K. Dutta, A. Sarkar, N. Das, and K. Roy, "A survey on different feature extraction methods for writer identification and verification," International Journal of Applied Pattern Recognition, vol. 7, no. 2, pp. 122-144, 2023.

[2] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," Journal of Applied Science and Technology Trends, vol. 1 no. 2, pp. 56-70, 2020.

[3] R. Kaur, M. Uppal, and D. Gupta, "A comprehensive and comparative study of handwriting recognition system," in Proc. of 2023 IEEE Renewable Energy and Sustainable E-Mobility Conference (RESEM), 2023, pp. 1-6.

[4] H. Freeman, "Techniques for the digital computer analysis of chain-encoded arbitrary plane curves," in Proc. of Electron, 1961, pp. 412-432.

[5] M. Kadi, Y. Douzi, Y. and M. B. Nasri, "A generalized freeman chain code for offline arabic character recognition," in Proc. of International Conference on Electronic Engineering and Renewable Energy Systems 2022, 2022, pp. 47-55.

[6] J. T. Wu, and J. J. Ding, "Improved angle freeman chain code using improved adaptive arithmetic coding," in Proc. of 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2020, pp. 181-184.

[7] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk, and W. Al-Atabany, "Honey badger algorithm: new metaheuristic algorithm for solving optimization problems," Mathematics and Computers in Simulation, vol. 192, pp. 84-110, 2022.

[8] R. Khajuria, S. Yelisetti, R. Lamba, and R. Kumar, "Optimal model parameter estimation and performance analysis of PEM electrolyzer using modified honey badger algorithm," International Journal of Hydrogen Energy, vol. 49, pp. 238-259, 2023.

[9] A. Fathy, H. Rezk, S. Ferahtia, R. M. Ghoniem, and R. Alkanhel, "An efficient honey badger algorithm for scheduling the microgrid energy management," Energy Reports, vol. 9, pp. 2058-2074, 2023.

[10] S. N. Qasem, "A novel honey badger algorithm with multilayer perceptron for predicting COVID-19 time series data," The Journal of Supercomputing, vol. 80, no. 3, pp. 3943-3969, 2024.

[11] A. A. Engkamat, "Enhancement of parallel thinning algorithm for handwritten characters using neural network," MSc. Thesis, UTM, Malaysia, 2015.