The Heuristic Extraction Algorithms for Freeman Chain Code of Handwritten Character

Dewi Nasien dwien82@gmail.com
Faculty of Computer Science and Information System (FSKSM)
Universiti Teknologi Malaysia
Skudai, 81310, Malaysia

Habibollah Haron habib@utm.com
Faculty of Computer Science and Information System (FSKSM)
Universiti Teknologi Malaysia
Skudai, 81310, Malaysia

Siti Sophiayati Yuhaniz sophia@utm.com
Faculty of Computer Science and Information System (FSKSM)
Universiti Teknologi Malaysia
Skudai, 81310, Malaysia

Abstract

Handwriting character recognition (HCR) is the ability of a computer to receive and interpret handwritten input. In HCR, there are many representation schemes and one of them is Freeman's chain code (FCC). Chain code is a sequence of code directions of a character and connection to a starting point which is often used in image processing. The main problem in representing a character using FCC is that the results of the extractions depend on the starting points. Unfortunately, the studies about FCC extraction using one continuous route and to minimize the length of chain code to FCC from a thinned binary image (TBI) have not been widely explored. To solve this problem, heuristic algorithms are proposed to extract the FCC to correctly represent the characters. This paper proposes two heuristics algorithms that are based on randomized and enumeration algorithms to solve the problems. As a problem solving technique, the randomized-based algorithm makes the random choices while the enumeration-based algorithm enumerates all possible candidates for a solution. The performance measurements of the algorithms are the route lengths and computation times. The experiments on the algorithms are performed based on the chain code representation derived from established previous works of Center of Excellence for Document Analysis and Recognition (CEDAR) dataset which consists of 126 upper-case letter characters. The experimental results show that the route lengths for both algorithms are similar but the computation times of the enumeration-based algorithm is higher than the randomized-based algorithm. This is because the enumeration-based algorithm considers all branches in route walk.

Keywords: Handwritten Character Recognition, Freeman Chain Code, Heuristic Algorithm, The Randomized-Based Algorithm, The Enumeration-Based Algorithm.

1. INTRODUCTION

Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input such as digital cameras and other devices. The goal of handwriting is to identify input characters or image correctly and then to analyze them. Basically, the handwritten characters are varied from one person to another person. Due to this wide range of variability, it is difficult to recognize a handwritten character by a machine. Comparing the achievements of various researches in this field is quite difficult as the databases and general approaches might differ [1]. For instance, different databases that are used to extract the characters will give different results and different complexity. It is also the same as for the general approaches which is found in various kinds of handwritten character such as digit, numeral, cursive script, symbols, and scripts, including English, Tamil, Bangla, Arabic, Chinese, Latin and others.

Handwritten character recognition (HCR) can be defined into two components, namely off-line and on-line recognition. Off-line recognition is a system that accepts its input from a digital scanner using the image processing algorithm. On the other hand, on-line recognition system accepts input data from on-line input devices and then computes the relationships between points to extract the features in real time. Off-line characters are less accurate than the on-line system. However, off-line character recognition has been extensively studied over the last few decades and many commercial systems are available today. Some of its application areas are automatic postal sorting, bank check processing and form processing.

HCR can be divided into three stages namely pre-processing, feature extraction and classification, which are the actual recognition process. Pre-processing stage is to produce a clean character image that can be used directly and efficiently by the feature extraction stage. Feature extraction stage is to remove redundancy from data, producing a feature vector and can be used by the classification. Lastly, classification stage is to recognize characters or words.

This paper only concentrates on feature extraction stage of a HCR. Feature extraction stage in HCR is a very important field of the image processing and object recognition system because it is used by classifier to classify the data. The basic task of feature extraction and selection is to find out a group of the most effective features for classification that is by compressing from high-dimensional feature space to low-dimensional feature space, to design classifier effectively [2]. Fundamental component of feature extraction is called features. As in many practical problems, it is often not easy to find those with most effective features [3]. This makes features extraction and selection as one of the most difficult and challenging tasks in pattern recognition system, data mining, and other fields. So, the task of human expert is to select features that allow effective and efficient recognition of pattern. The abundant with features makes it too difficult to calculate, because it will occupy a lot of memory space and increase the computerization time. On the other hand, many features will include many correlation factors respectively, which results to information repeat and waste [4].

Therefore, we must take certain measures to decrease the feature dimension but without decreasing the recognition effect. This is called the problems of a feature optimum extraction or selection [5-7] (cite in [4]). Types of features are depended on the system in which they are implemented. One of them is a shape feature. There are two procedures for extracting these features. The first procedure is boundary extraction that based on the outer boundary shape. The second procedure is region extraction that works with the whole shape as an object.

Freeman's chain code (FCC) is one of the techniques representations based on the boundary extraction which useful for image processing, shape analysis and pattern recognition. Chain code representation gives a boundary of character image where those codes represent the direction of where is the location of the next pixel. Unfortunately, the study about FCC construction using one continuous route and minimizing the length of chain code to FCC from a thinned binary image (TBI) has not been widely explored. To solve this problem, heuristic methods are used to extract the FCC that is correctly representing the characters. The success rate in recognizing character depends on all stages in HCR, which are pre-processing, feature extraction and classification.

Heuristic algorithm is used in extracting a feature of isolated Latin character recognition. Latin characters, especially the upper-case characters are usually having branches that make the traversing process difficult. There are two kinds of the heuristics algorithm to solve the HCR problems, which are the randomized-based and the enumeration-based algorithms. The randomized-based algorithm is a problem solving technique with makes the random choices while the enumeration-based algorithm is a problem solving technique with enumerating all possible candidates for a solution.

This paper is organized as follows. Section 2 presents the chain code scheme in HCR. Section 3 describes the experimental design and pre-processing stage. Section 4 describes the heuristic extraction algorithms proposed, which are the randomized-based and the enumeration-based algorithms. Section 5 presents the parameter value setting. Section 6 describes the experimental result and discussion and followed by a conclusion in Section 7.

2. CHAIN CODE SCHEMA IN HCR

Chain code is a representation technique of different model shapes. Chain code is used for descriptions of object borders in image processing. The first approach of chain code was introduced by Freeman in 1961 that is known as Freeman Chain Code (FCC) [8]. There are many kinds of chain code algorithms, which have been developed through extension of FCC and enhancement of chain code, which are Papert, Primitives Chain Code (PCC), Vertex Chain Code (VCC), 3OT, Extended Vertex Chain Code (E_VCC), Variable Length Vertex Chain Code (VL_VCC) and Compressed Vertex Chain Code (C_VCC). Previous work in the literature about chain code representation can be found in [9-11].

There are two directions of chain code, namely 4-neighborhood and 8-neighborhood. This paper utilizes 8-neighbourhood in extraction of characters as shown in Fig. 1 (b). Usually researchers in other research works would start from zero until seven in 8-neighbourhood but for this paper will start from one until eight because is easy to distinguish the direction or non direction (value is zero) of chain code. The challenge of the chain-coding process would be very much on the way of the image would be traversing and the starting point of the traversing method [1]. A start point of a character will produce a different chain code direction even though is the same image. Randomly, the start point in a character is selected and then the best solution is searched.

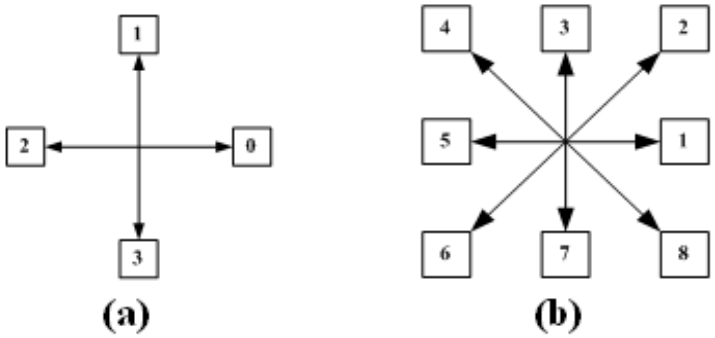


FIGURE 1: Two Directions of FCC : (a) 4-Neighborhood, (b) 8-Neighborhood.

Heuristic method is used for minimizing the length of the chain code. They are a the randomized-based and the enumeration-based algorithms. The use of them is motivated by following considerations:

- The starting node for the FCC construction influences its length. In addition, a handwritten character often has several branches, and this make it is difficult to decide where it should go. Moreover, a revisit to the previous visited nodes is sometimes needed to visit all the nodes. These difficulties motivate the use of heuristic approaches to solve it.

- The performance of the recognition stage depends on the data input provided from the previous stage. The input data for the recognition stage (in this case is the FCC) must correctly represent and distinguish each character. Since a handwritten character can be represented with several FCCs, the methods must have the ability to provide such results. In order to achieve this, heuristics is used to generate the FCC. Besides that, several adjustments are performed to the proposed heuristics to produce good FCC results.

TBI of pre-processing stage is used in the feature extraction for extracting the character recognition. The purpose of thinning is to delete the redundant information and at the same time retain the characteristic features of the character. Thinning is applied to find a skeleton of a character. Skeleton is an output of thinning process.

A number of papers have been proposed in chain code algorithms in HCR. The application of chain code approaches that have been used in several HCR problems such as a fast handwritten word recognition system for real time applications are presented in [12]. A new type of generalized chain code (GCC) [13] is proposed for lossless encoding of handwriting data captured by a tablet and real-time recognition of isolated on-line handwritten characters. An approach for fast computation of moments based on the 8-neighbour chain code is proposed in [14]. A discuss the issues of determination of upper and lower contours of the word, determination of significant local extrema on the contour and determination of reference lanes from contour representations of handwritten words [15]. A recognition method based on handwriting acceleration, line crossing points segmentation, macro structures (isolated traces), chain coding and time-frequency analysis [16]. An efficient technique for recognizing closed shapes is given [17]. The result shows that the calculation of the proposed algorithm is simple as well as efficient in time and space complexity.

Besides that, a technique for the exchange of handwritten information through telecommunications is proposed [18]. Freeman's chain code is used to discriminate sets of feature vectors in a multi-font Bengali character recognition system [19]. A new fast method for line segment extraction from images and utilized the method for Persian signature recognition. The proposed Chain Based Line Segment Extraction (CBLSE) method is utilizing a chain code representation of edge points, and split is the chains in higher curvature points [20]. A novel approach of fingerprint recognition based on Flow Pattern, and Chain Coded String Matching technique [21]. A robust and effective of an HCR system based on modified chain code of Persian and Arabic alphabets is proposed in [22]. In [23], a HCR system is proposed by finding the contour of a handwriting image and a set of chain code from the global and local histogram. A method of extracting chain codes of a thinned binary image using Differential Evolution (DE) and Particle Swarm Optimization (PSO) to find a continuous route which covers all the nodes of the image [24]. A recognition model for English handwritten (lowercase, uppercase and letter) character recognition that uses Freeman's chain code (FCC) as the representation technique of an image character [25]. Scale invariant code is a code that will remain identical even when to change the scale of an object. The output chain code will remain same, so it can be used in object representation and recognition. The algorithm is developed for Nastaliq characters [26].

3. CHAIN CODE SCHEMA IN HCR

This section explains the experimental design and pre-processing stage of HCR. Firstly, by using input character that is digitized, thinning is performed as pre-processing stage in the HCR. The output of TBI is a skeleton that is used in feature extraction stage. The example of original image and its skeleton of character is shown in Fig. 2. From the first stage to the second stage, the representation technique of pattern is needed. In this case, FCC is selected and is used to represent the character. Moreover, feature extraction stage is used to extract the character using FCC. Finally, the desired output of chain code is obtained. This work applies thinning algorithm that is proposed by Engkamat [27] in extracting the FCC. The FCC is then analyzed using heuristic methods, which are the randomized-based and the enumeration-based algorithms. The proposed heuristic chain code extraction algorithm of HCR is shown in Fig. 3.

FIGURE 2: Original Image and Skeleton.

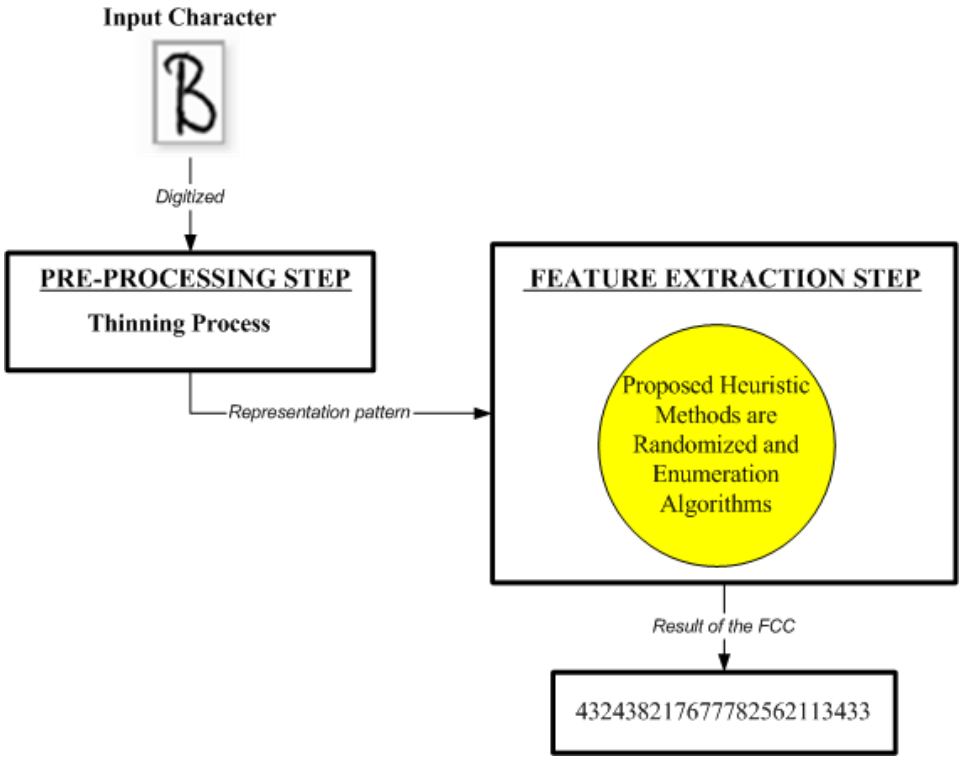


FIGURE 3: The Proposed Heuristic Chain Code Extraction Algorithm.

Pre-processing stage only involves the thinning process. TBI by Engkamat is used in feature extraction for extracting the character recognition. The dataset from CEDAR is used as the input data in this experiment. This paper only concentrates on 126 upper-case Latin characters (A-Z) as shown in Table 1. The pixel input of original CEDAR is 50x50 pixels and the output of TBI is similar too. Fig. 4 shows the example of the TBI obtained using the proposed thinning algorithm [27]. TBI is presented in a binary image. Binary image is a representation with only two possible gray values for each pixel, such as “0” and “1”. Background is represented by “0” and foreground is represented by “1”. Fig. 5 shows the coordinate (x, y) of TBI for B1 character. Fig. 5 will be used in the explanation of the proposed algorithm. The detail process of both algorithms is discussed in Section 4.

Characters	The Total Number for Every Character
Q	3 x 1
I, P	4 x 2
A - H J - O R - Z	5 x 23
TOTAL	126 Characters

TABLE 1: The Detail of Image Properties for Dataset.

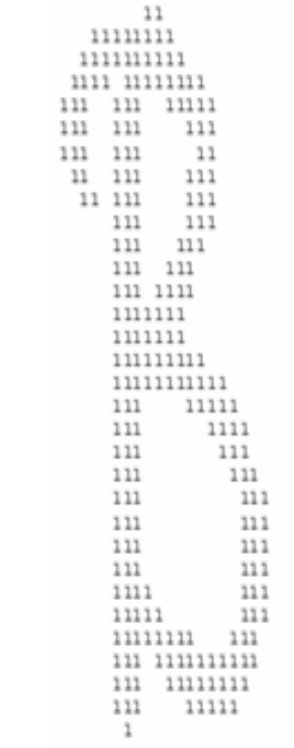
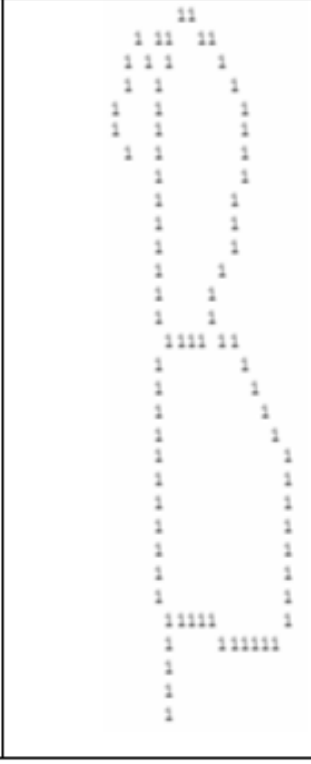
Original Image	TBI Proposed by Engkamat
	

FIGURE 4: The Example of TBI Special for B1 Character by Engkamat [27].

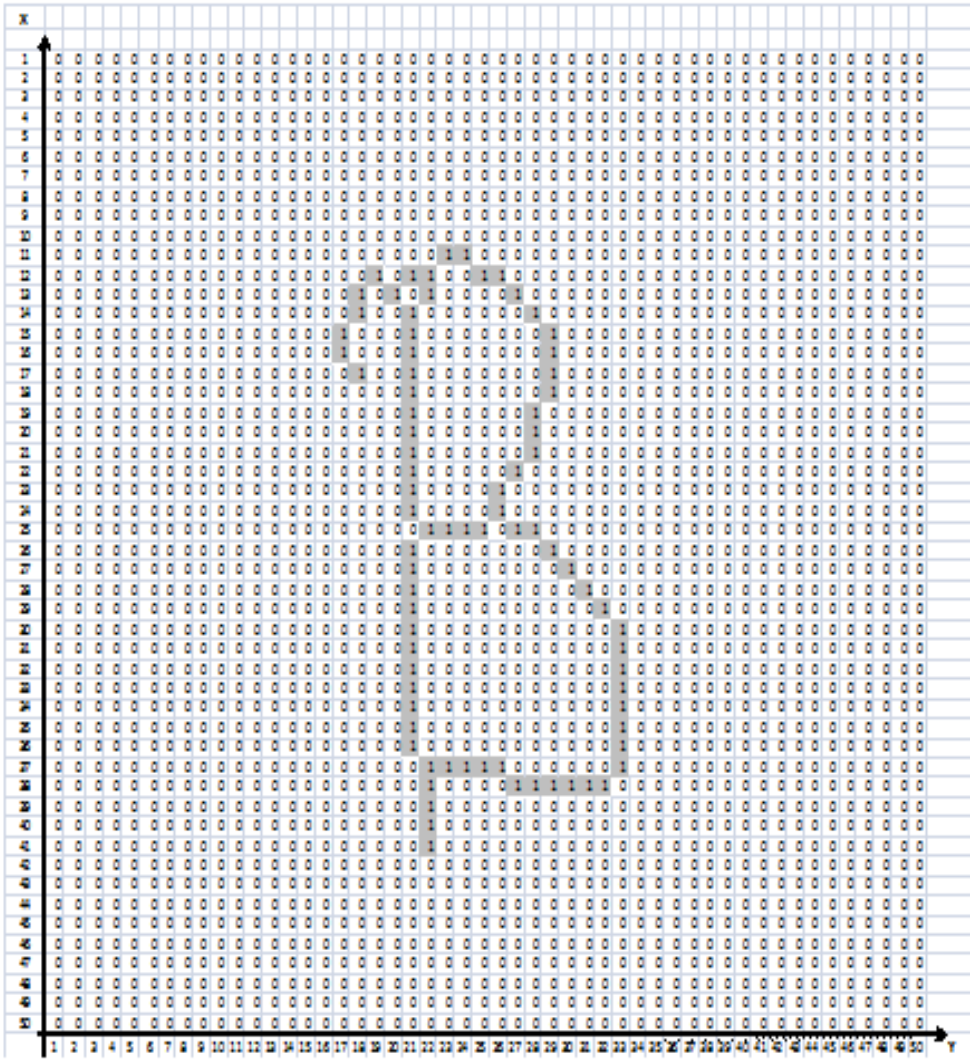


FIGURE 5: The Coordinate (X, Y) of TBI for B1Character.

4. THE HEURISTIC EXTRACTION ALGORITHMS

Computer and algorithm are used to solve problems where humans are not being able to accomplish. Heuristic method is used to solve the problem to generate FCC of a handwritten binary image. Heuristic methods are excellent in finding approximate solutions. Two important terms in this method are the acceptable time rate and space complexity. Time complexity refers to running time of the program while space complexity refers to the amount of computer memory required during the program execution.

The application of heuristic, especially for HCR has not widely explored by researchers. The example of heuristic in HCR is for segmentation for Arabic script [28], a neuro-heuristic for segmenting handwritten Arabic text [29], a novel method for analyzing conceptual design data using heuristics [30], segmenting Chinese's unknown words by the heuristic method [31] and Chinese's text segmentation using the statistical approach [32].

The proposed of algorithms is the similar study of application of heuristic was conducted by [33] to handwritten numeral recognition. Heuristic is used in the feature extraction. Before heuristic feature extraction, a thinning process is needed to thin the numeral image toward a skeleton

form. After thinning, the trace of the whole numeral skeleton is carried out, and heuristic features are extracted step by step. The extended chain code is selected as a technique of representation. Extended chain code has twenty four directions with 5 x 5 window for a resolution of 128 x 128 in the numeral image. An experiment is conducted for 400 randomly written numerals with a Hamamatzu C-100 camera as input device. The recognition rate is about 97%.

4.1 The randomized-based Algorithm

The randomized-based algorithm is an algorithm that makes random (pseudo-random) choices. Good the randomized algorithm is the one that comes with a very high probability of correct computed results [34]. The amount of time and space are needed to solve the computation of complexity theory. For more detail about the randomized can be seen in [35].

A complexity theory is a mathematical logic of the computational problem which is a certain computational resource. Polynomial (P) and Non-deterministic Polynomial (NP) are parts of complexity of computational theory, which are to solve a given problem. The P of class complexity that can be solved by a deterministic machine in polynomial time while NP that can be solved by a non-deterministic machine in polynomial time. Boolean satisfiability problem, hamiltonian path problem and vertex cover problem are problems that can be solved effectively in NP. All the problems in this class have the property that their solutions can be checked efficiently [36]. This paper describes a randomized-based algorithm using the concept of Hamiltonian path.

The following describes the randomized-based algorithm in more detail. The explanation is referred to the TBI that is used as the input to the algorithm as shown in Fig. 5.

1. Initialize data
    a. Choose the first node
        - To find all the nodes of the image, number of nodes, size of the image matrix, to find the coordinate of node such as left upper, right upper, left lower and right lower.
    b. Iteration for end node
        - To find all the end nodes and the number of the end nodes.
    c. Choose neighbours with examine neighbour image
    d. Set the output parameter
        - The number of maximum iteration, FCC, X and Y first node output, route length, maximum display and maximum route.
    e. Set the iteration parameter
        - X and Y neighbour unvisited, FCC unvisited, X and Y neighbour visited, FCC visited, image matrix, X and Y new route, and new FCC.
2. Find the route the randomized-based randomly
    a. Identify node-method and end-node-method of each node started from the first node.
        - Node-method is to find the first character for every aspect boundary such as left upper, right upper, left lower and right lower.
        - End-node-method is to find the first character based on the end node position of a character.
        - Example: For B1 character in Fig. 5, the position of the node-method is left upper (12,19), right upper (12,26), left lower (41,22) and right lower (37,33). The end-node-method consist of two end-node-methods that are (17,18) and (41,22).
    b. If the number of visited node is less than the number of nodes, identify three types of characteristics, which are unvisited, visited and taboo neighbours.
        - Unvisited neighbours are nodes that never went through the route searching.
        - Visited neighbours indicate the nodes that have gone through the route searching.
        - Taboo neighbours are used to keep track of the visited search space and revisited node with one step before the current node.
    c. If unvisited, visited or taboo neighbours then select one node randomly.
3. It selects the best solution that will survive for the next replication.
4. Algorithm stops at the pre-determined number of replications

Based on Fig. 5, the FCC length for B1 is shown in Table 2. The pseudo-code of the randomized-based algorithm is shown in Table 3 while the description of the randomized-based algorithm in Fig. 6.

4	3	2	3	2	8	2	1	7	6	7	7	7	7	...	...	...	...	...	...	4	5	4	5	0	0
4	3	2	3	2	8	8	7	7	7	7	7	7	7	...	...	...	...	...	...	5	6	7	4	0	0
4	3	2	3	2	8	2	8	3	2	1	8	1	8	...	...	...	...	...	...	3	3	3	3	0	0
.																									.
.																									.
.																									.
.																									.
4	3	2	3	2	8	8	7	7	7	7	7	7	7	...	...	...	...	...	...	4	5	6	7	4	0
4	3	2	3	2	8	2	8	6	7	7	7	7	7	...	...	...	...	...	...	4	5	4	5	6	0
4	3	2	3	2	8	8	7	7	7	7	7	7	7	...	...	...	...	...	...	4	5	6	5	8	0

→ 10 x 90

TABLE 2: The FCC (Part of the Output) for B1 Character Using the Randomized-Based Algorithm Into Matrix.

```
Initialize Data
while Termination Not Met do
    Select First Node Randomly
    {Node-Method, End-Node-Method}
        while Number of Visited Node < Number of Node do
            if there are Unvisited Neighbours
                Select One Node Randomly
            elseif there are Visited Neighbours
                Select One Node Randomly
            elseif there are Taboo Neighbours
                Select One Node Randomly
            endif
        endwhile
endwhile
Display Solutions
```

TABLE 3: The Pseudocode of the Randomized-Based Algorithm.

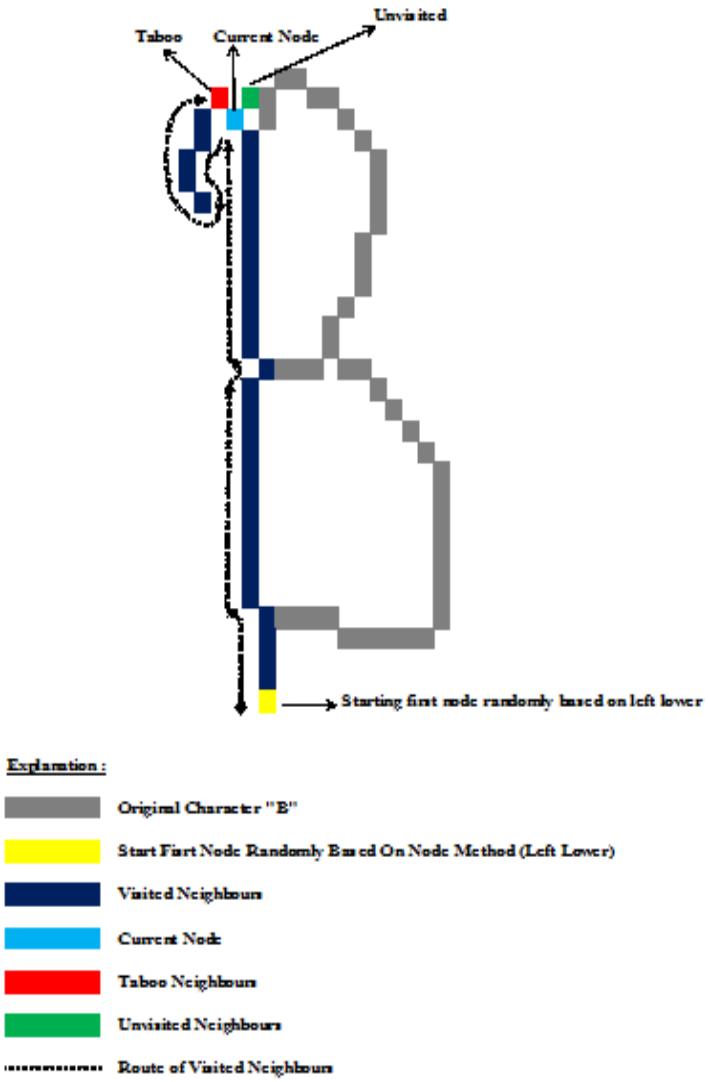


FIGURE 6: Description of the Randomized-Based Algorithm.

4.2 The Enumeration-Based Algorithm

The enumeration-based search or brute-force search or exhaustive search is a problem solving technique by enumerating all possible candidates for the solution. This means it exhaustively search through all possible combinations until a solution is found. We will use the term the enumeration-based for this paper.

The enumeration-based is simple to implement. It is used to find a solution if it exists. In the enumeration-based search, the problem size is limited to reduce the set of candidate solutions to a manageable size. The capability of computer is important where it influents the result of the enumeration-based algorithm. When the capability of computer is a not high enough, the enumeration-based algorithm needs a longer time. The quantity and time of the enumeration-based can be to reduce to enhance the performance of computer capability to an acceptable scope.

The following describes the enumeration-based algorithm in more detail. The explanation is referred to the TBI shown in Fig. 5 as input to the algorithm.

1. Initialize data

a. Set the output parameter
   - To find the all the node in character, number of nodes, maximum route, maximum display, allocation, size of image matrix, maximum iteration, FCC, X & Y first node and route length
b. Choose neighbours with examines neighbour images
c. Assign the first point
   - To find the coordinate of node method and end node method
d. Set the iteration parameter
   - Number visited, X and Y neighbour unvisited, FCC unvisited, X and Y neighbour visited, FCC visited
e. Sign the point has already used

2. Find the route the enumeration-based randomly
   a. Identify node-method and end-node-method of each node started from the first node.
      - Node-method is to find the first character for every aspect boundary such as left upper, right upper, left lower and right lower.
      - End-node-method is to find the first character based on the end node position of a character.
      - Example: For B1character in Fig. 5, the position of the node-method is left upper (12,19), right upper (12,26), left lower (41,22) and right lower (37,33). The end-node-method consist of two end-node-methods that are (17,18) and (41,22).
   b. If the number of visited node less than the number of nodes, identify three types of characteristics, which are unvisited, visited and taboo neighbours.
      - Unvisited neighbours are nodes that never went through the route searching.
      - Visited neighbours indicate the nodes that have gone through the route searching.
      - Taboo neighbours are used to keep track of the visited search space and revisited node with one step before the current node.
   c. If unvisited, visited or taboo neighbours then select branch the walk to the unvisited neighbours.
3. It selects the best solution that will survive for the next replication.
4. Algorithm stops at the pre-determined number of replications.

Based on the Fig. 5, the undefined FCC length for B1. This problem appears because there are many branches that must be traversed but the capability of memory computer is not supported. For “B2-B5” in B character the result and FCC length is produced. The pseudo-code of the enumeration-based algorithm is as shown in Table 4 while the description of the enumeration-based algorithm in Fig. 7.

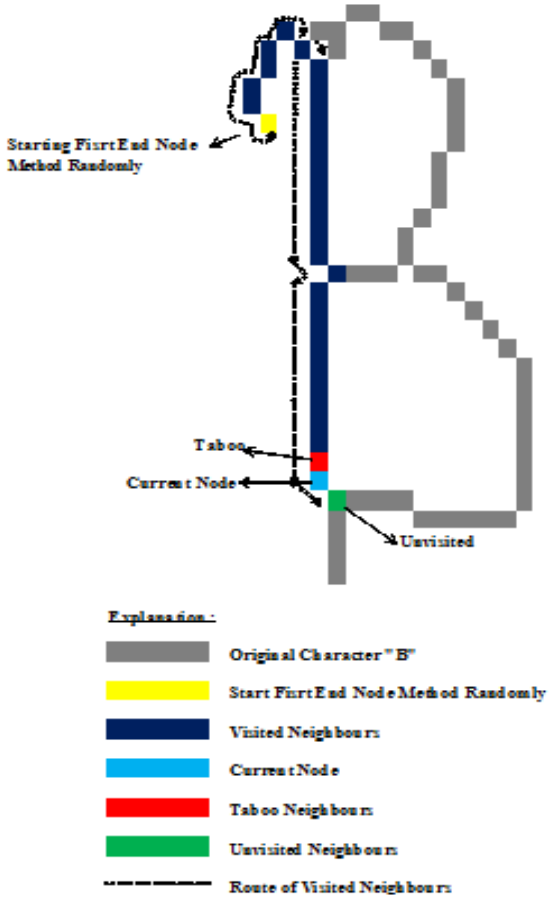


FIGURE 7: Description of Enumeration Algorithm.

```
Initialize Data
while Termination Not Met do
    Start the FCC from Node-Method and End-Node-Method
        while Number of Visited Node < Number of Node do
            if there are Unvisited Neighbours
                Branch the Walk to the Unvisited Neighbours
            elseif there are Visited Neighbours
                Branch the Walk to the Unvisited Neighbours
            elseif there are Taboo Neighbours
                Branch the Walk to the Unvisited Neighbours
            endif
        endwhile
endwhile
Display Solutions
```

TABLE 4: The Pseudocode of the Enumeration-Based Algorithm.

5. PARAMETER VALUE SETTING

The performances for both methods are evaluated. The criteria termination of both proposed methods are shown below:

1. The experiment with five replications, each replication consists of ten best of FCC solutions.
2. The maximum iteration is two thousand five hundred iterations where that size of image matrix row * size of image matrix column (50 pixel of image matrix row * 50 pixel of image matrix column = 2500 iterations)
3. Maximum route is 2 * number of node where that the number of node is the length of X all of the nodes.
4. For enumeration-based method, the allocation memory is one hundred. Memory allocation is important in order to avoid frequently changing the length of the variable when it is executed.

The specification of hardware and software are shown below:

1. Notebook Compaq Presario CQ40
   - Memory (RAM) 3GB
   - Processor is AMD Turion™X2 Dual Core Mobiles RM-72 2.1GHz
2. Matlab R2008a (version 7.6)

6. EXPERIMENTAL RESULT AND DISCUSSION

This section describes the experimental results of the experiment. These two algorithms are used to generate the continuous FCC which acts as the image features. The performance of both proposed methods are route lengths (number of chain code length) and computation time (seconds). After that the results are compared to conclude which that has to perform better.

Based on Table 1, the route length and computation time for B characters ("B1-B5") of both algorithms are shown in Table 5. After all B characters ("B1-B5") are finished in terms of route length and computation time, the next step is to classify the route lengths into three categories: best, average and worst. Best is the minimum value group of the number of iterations in a set of values. Average is calculated by adding a group of numbers of iterations and then dividing by the count numbers of every character. Worst is the maximum value a group of numbers of an iteration in a set of values. Furthermore, for the computation time categories are average and total. Average is calculated by adding a group of number iterations and then divided by the count numbers of every character. Total is the sum of the group of number iterations in a set of values.
.

Char	Heuristic Methods	Route Length (Number of chain code length)					Computation Time (Seconds)				
		R1	R2	R3	R4	R5	R1	R2	R3	R4	R5
B1	Random	89	89	89	89	89	1.501	1.391	1.412	1.421	1.423
	Enumeration	-	-	-	-	-	-	-	-	-	-
B2	Random	72	72	72	72	72	1.051	1.129	1.124	1.112	1.06
	Enumeration	72	72	72	72	72	1.585	1.281	1.069	1.24	1.267
B3	Random	104	104	104	104	104	1.9	1.967	2.006	1.964	1.894
	Enumeration	104	104	104	104	104	1.275	1.203	1.242	1.233	1.229
B4	Random	73.2	73.2	73.2	73.2	73.2	1.239	1.196	1.265	1.266	1.204
	Enumeration	73.2	73.2	73.2	73.2	73.2	1.136	1.204	1.18	1.199	1.315
B5	Random	104.11	104.11	104.11	104.11	104.11	1.898	1.754	1.807	1.756	1.779
	Enumeration	104.11	104.11	104.11	104.11	104.11	1.122	1.155	1.127	1.178	1.167

TABLE 5: Character of "B1-B5" with Five Replications

Table 6 shows the characters of "B1-B5" for route length in terms of best, average and worst. On the other hand, the computation time in terms of average and total. The results for B1 are omitted due to lack of B1 length from the enumeration-based algorithm. Furthermore, the result of B1 under randomized is ignored too for easy to calculate route length and computation time for only B character as shown in Table 7. The detail result of both algorithms is as shown in Table 8. The fastest and slowest of computation times for both algorithms are shown in Table 9 and Figure 8

Char	Heuristic Methods	Route Length (Number of chain code length)			Computation Time (seconds)	
		Best	Average	Worst	Average	Total
B1	Random	-	-	-	-	-
	Enumeration	-	-	-	-	-
B2	Random	72.00	72.00	72.00	1.10	5.48
	Enumeration	72.00	72.00	72.00	1.29	6.44
B3	Random	104.00	104.00	104.00	1.95	9.73
	Enumeration	104.00	104.00	104.00	1.24	6.18
B4	Random	73.20	73.20	73.20	1.23	6.17
	Enumeration	73.20	73.20	73.20	1.21	6.03
B5	Random	104.11	104.11	104.11	1.80	9.00
	Enumeration	104.11	104.11	104.11	1.15	5.75

TABLE 6: Characters of "B1-B5" Using Route Length and Computation Time with Five Replications.

Char	Heuristic Methods	Route Length (Number of chain code length)			Computation Time (seconds)	
		Best	Average	Worst	Average	Total
B	Random	88.33	88.33	88.33	1.52	7.59
	Enumeration	88.33	88.33	88.33	1.22	6.10

TABLE 7: The Route Length and Computation Time for All of B Character.

Char	Methods	Route Length (Number of chain code length)			Computation Time (seconds)	
		Best	Average	Worst	Average	Total
A	Random	80.48	80.50	80.55	1.50	30.06
	Enumeration	80.48	80.48	80.48	4.77	95.4
B	Random	88.33	88.33	88.33	1.52	30.37
	Enumeration	88.33	88.33	88.33	1.22	24.41
C	Random	74.09	74.09	74.09	0.93	18.63
	Enumeration	74.09	74.09	74.09	1.22	24.42
D	Random	80.52	80.52	80.52	1.24	30.96
	Enumeration	80.52	80.52	80.52	1.17	29.35
E	Random	62.17	62.17	62.17	1.01	15.1
	Enumeration	62.17	62.17	62.17	1.16	17.41
F	Random	65.46	65.46	65.46	1.09	27.16
	Enumeration	65.46	65.46	65.46	1.15	29.35
G	Random	85.00	85.00	85.00	1.13	5.64
	Enumeration	85.00	85.00	85.00	1.47	7.35
H	Random	85.73	85.73	85.73	1.44	28.9
	Enumeration	85.73	85.73	85.73	1.27	25.5
I	Random	54.88	54.88	54.88	0.96	19.21
	Enumeration	54.88	54.88	54.88	1.31	25.04
J	Random	67.04	67.04	67.04	1.12	28.03
	Enumeration	67.04	67.04	67.04	1.17	29.22
K	Random	64.25	64.25	64.25	1.12	22.34
	Enumeration	64.25	64.25	64.25	1.80	36.01
L	Random	46.05	46.05	46.05	0.67	16.86
	Enumeration	46.05	46.05	46.05	1.19	29.87
M	Random	75.17	75.17	75.17	1.14	17.06
	Enumeration	75.17	75.17	75.17	1.15	17.26
N	Random	60.68	60.68	60.68	0.94	23.4
	Enumeration	60.68	60.68	60.68	1.19	29.77
O	Random	73.64	73.64	73.64	1.17	23.48
	Enumeration	73.64	73.64	73.64	1.26	25.21
P	Random	64.99	64.99	64.99	1.05	15.69
	Enumeration	64.99	64.99	64.99	1.15	17.27
Q	Random	84.75	84.75	84.75	1.53	15.3
	Enumeration	84.75	84.75	84.75	1.29	12.86
R	Random	80.87	80.87	80.87	1.37	34.3
	Enumeration	80.87	80.87	80.87	1.20	29.92
S	Random	76.87	76.87	76.87	1.12	16.77
	Enumeration	76.87	76.87	76.87	1.48	22.13
T	Random	58.45	58.45	58.45	0.92	18.34
	Enumeration	58.45	58.45	58.45	1.37	27.44
U	Random	78.25	78.25	78.25	1.05	26.17
	Enumeration	78.25	78.25	78.25	1.23	30.76
V	Random	65.80	65.80	65.80	0.81	20.14
	Enumeration	65.80	65.80	65.80	1.18	29.6
W	Random	65.47	65.47	65.47	1.22	30.44
	Enumeration	65.47	65.47	65.47	2.10	52.43
X	Random	56.78	56.78	56.78	0.99	19.83
	Enumeration	56.78	56.78	56.78	0.88	23.72
Y	Random	49.93	49.93	49.93	0.79	18.9
	Enumeration	49.93	49.93	49.93	1.20	29.89
Z	Random	61.37	61.37	61.37	0.95	19.08
	Enumeration	61.37	61.37	61.37	1.21	24.17

TABLE 8: Results of the Proposed Algorithms.

The Fastest of Computation Time (seconds)				The Lowest of Computation Time (seconds)			
Char	Methods	Average	Total	Char	Methods	Average	Total
R	Randomized	1.37	34.30	G	Randomized	1.13	5.64
A	Enumeration	4.77	95.40	G	Enumeration	1.47	7.35

TABLE 9: The Fastest and Slowest of The Computation Time of Both Algorithms.

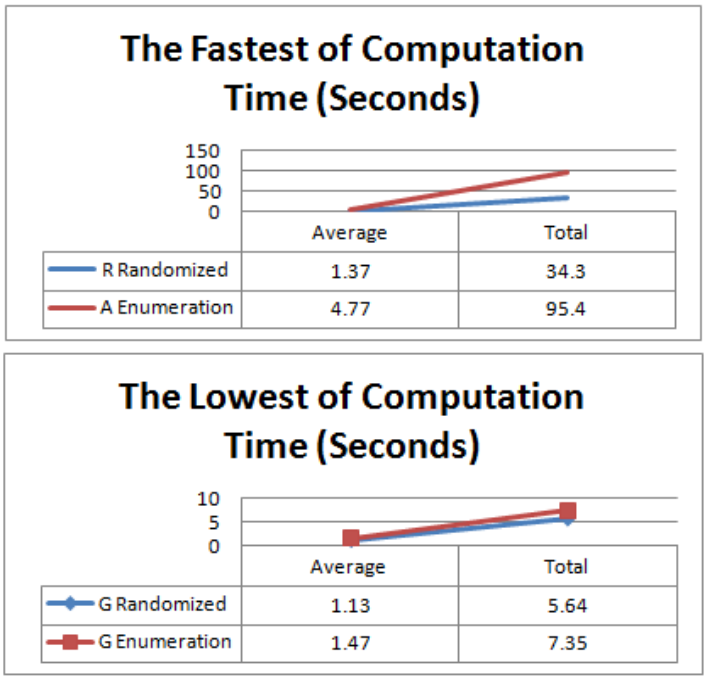


TABLE 10: The graph of Fastest and Slowest of The Computation Time of Both Algorithms.

The main advantage of the enumeration-based algorithm is that the algorithm considers all the branches that exist in the node of the characters. It made three routes walk namely, unvisited, visited and taboo. The weakness of this algorithm is that it could not recognize all characters as it takes a lot of computation time and space.

The randomized-based algorithm gives better performance as during the experiment it takes 2500 iterations that are used to prioritize the heuristic rules of searching the unvisited, visited and taboo. However, it does not guarantee to produce the best solution, for example, character A.

The route lengths for the proposed algorithms are shown in Table 8. It can be seen that both approaches give similar results in term of the resulting route lengths. However, in term of computation times, the randomized-based algorithm is faster than the enumeration-based algorithm. The proposed randomized-based algorithm takes 572.16 seconds. Meanwhile, the proposed enumeration-based algorithm takes 745.76 seconds.

This research work has demonstrated the application of Freeman Chain Code in the problem of HCR. In the HCR application, the FCC is extracted as the features for the FCC classifier. However, as with many other chain coding problems, finding the best routes with minimal route length is the common problem. In this research work, heuristic methods have been applied when extracting FCC from handwritten characters. The experimental results show that the proposed extraction algorithms using the two heuristic methods have a high success rate of 83%.

7. CONSLUSION

Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input. Character recognition is a subset of handwriting recognition. There are many kinds of handwritten character and one of them is Latin characters. Latin characters consist of 26 characters which are upper-case letter (A-Z) and lower-case letter (a-z).

HCR can be divided into three stages namely pre-processing, feature extraction and classification, which are the actual recognition process. This paper only concentrates on pre-processing and feature extraction stages of a HCR. Pre-processing stage is to produce a clean character image that can be used directly and efficiently by the feature extraction stage. In addition, feature extraction stage is to remove redundancy from data, producing a feature vector and can be used by the classification. From the first stage to the second stage, the representation technique of pattern is needed. In this case, FCC is selected and is used to represent the character.

FCC is one of the techniques representations based on the boundary extraction. Unfortunately, the study about FCC construction using one continuous route and minimizing the length of chain code to FCC from a TBI has not been widely explored. To solve this problem, heuristic methods are used to extract the FCC that is correctly representing the characters. Heuristic method is selected for minimizing the length of the chain code. They are the randomized-based and the enumeration-based algorithms. The use of them is motivated by following considerations:

- The starting node for the FCC construction influences its length. In addition, a handwritten character often has several branches, and this make difficult to decide where it should go. Moreover, a revisit to the previous visited nodes is sometimes needed to visit all the nodes. These difficulties motivate the use of heuristic approaches to solve it.
- The performance of the recognition stage depends on the data input provided from the previous stage. The input data for the recognition stage (in this case is the FCC) must correctly represent and distinguish each character. Since a handwritten character can be represented with several FCCs, the methods must have the ability to provide such results. In order to achieve this, heuristics is used to generate the FCC.

Moreover, both of the proposed algorithms try to find a collection of good FCC solution which minimize the FCC length. The route length (best, average and worst) and computation time (average and total) are selected in this experiment because the resulting FCC is depended on the starting point and automatically affected on the route length and how many time it is needed to solve the chain code. This method enables us to extract and recognize such difficult characters in relatively shorter computational time and route length.

The experiments on the algorithms are performed based on the chain code representation derived from CEDAR dataset which consists of 126 upper-case Latin characters. The experimental results show that the route lengths for both algorithms are similar. The randomized-based algorithm does not guarantee to produce the best solution, for example, character A. On the contrary, in the enumeration-based algorithm from 126 upper-case, there is 21 characters could not to solved. The randomized-based algorithm is faster than the enumeration-based algorithm. This is due to the enumeration-based algorithm that has to consider all branches in route length, which are unvisited, visited and taboo.

8. ACKNOWLEDGMENT

The authors honorably appreciate Ministry of Science, Technology and Innovation (MOSTI) for the eScienceFund and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the support in making this projects success, also to Komarudin in assisting the development and testing of the algorithm.

9. REFERENCES

1. Suliman, A. Shakil, A. Sulaiman, M. N. Othman, M. and Wirza, R. "*Hybrid of HMM and fuzzy logic for handwritten character recognition*". In Proceedings of Information Technology. Kuala Lumpur, Malaysia, 2008

2. Shi-Fei, D. Wei-Kuan, J. Chun-Yang, S. and Zhong-Zhi, S. "*Research of pattern feature extraction and selection".* In Proceedings of Machine Learning and Cybernetics. Kunming, China, 2008

3. Zhaoqi, B. and Xuegong, Z. "*Pattern Recognition"*, 2nd Edition, Tsinghua University Press., (2000)

4. Shi-Fei, D. Zhong-Zhi, S. Vun-Cheng, W. and Shu-Shan, L. "*A novel feature extraction algorithm*". In Proceedings of Machine Learning and Cybernetics. Guangzhou, China, 2005

5. Jin, F. X. Wang, T. X. and Du, Z. X. "*Pattern recognition on surveying data information analysis*". Nonferrous Metal, 9(2):425-430, 1999

6. Ding, S. F. and Jin, F. X. "*Information features K-L transform based on information entropy*". Transactions Nonferrous Metals Society, 13(3):729-734. 2003

7. Turk, M. and Pentland, A. "*Face recognition using eigenfaces*". In Proceedings of Computer Vision and Pattern Recognition. Maui, HI , USA, 1991

8. Freeman, H. "*Techniques for the digital computer analysis of chain-encoded arbitrary plane curves*". In Proceedings of Electron. 1961

9. Liu, Y. K. and Zalik, B. "*An efficient chain code with huffman coding*". Pattern Recognition, 38(4):553-557, 2005

10. Sánchez-Cruz, H. Bribiesca, E. and Rodríguez-Dagnino, R. M. "*Efficiency of chain code to represent binary objects".* Pattern Recognition, 40(6):1660-1674, 2007

11. Wulandhari, L. A. and Haron, H. "*The evolution and trend of chain code scheme*". Graphics, Vision and Image Processing, 8(3):17-23, 2008

12. Gyeonghwan, K. and Govindaraju, V. "*Handwritten word recognition for real-time applications".* In Proceedings of the 3rd Document Analysis and Recognition. Montreal, Que. Canada, 1995

13. Hung, Y. "*A chain coding approach for real-time recognition of on-line handwritten characters*". In Proceedings of Acoustics, Speech and Signal Processing. Atlanta, GA , USA, 1996

14. Guang-Rong, J. Guo-Yu, W. Houkes, Z. Bing, Z. and Yan-Ping, H. "*A new method for fast computation of moments based on 8-neighbor chain code applied to 2-D object recognition"*. In Proceedings of Intelligent Processing System. Beijing, China, 1997

15. Madhvanath, S. and Govindaraju, V. "*Contour-based image preprocessing for holistic handwritten word recognition".* In Proceedings of the 4th Document Analysis and Recognition. Ulm, Germany, 1997.

16. Martinez, J. C. de J. Lopez, J. and Luna Rosas, F. J. "*A low-cost system for signature recognition.* In Proceedings of the 42nd Circuit and System. Las Cruces, NM , USA, 1999

17. Khalifa, I. H. Fahmi, M. S. H. Hassanien, A. E. and Elsalamony, H. A. R. M. "*Shape signature for recognition process*". In Proceedings of the 7th Radio Science. Minufiya, Egypt, 2000

18. Andrieux, J. and Seni, G. "*Coding efficiency of multi-ring and single-ring differential chain coding for telewriting application"*. Vision, Image and Signal Processing, 148(4):241-247, 2001

19. Mahmud, J. U. Raihan, M. F. and Rahman, C. M. *"A complete OCR system for continuous Bengali characters".* In Proceedings of Convergent Technologies. Asia-Pacific Region, 2003

20. Chalechale, A. and Mertins, A. "*Line segment distribution of sketches for Persian signature recognition".* In Proceedings of Convergent Technologies. Asia-Pacific Region*,* 2003

21. Paul, R. Nasif, M. S. and Farhad, S. M. "*Fingerprint recognition by chain coded string matching techniques".* In Proceeding of Information and Communication Technology. Dhaka, Bangladesh, 2007

22. Alaei, A. Pal, U. and Nagabhushan, P. "*Using modified contour features and svm based classifier for the recognition of persian/arabic handwritten numerals".* In Proceedings of the 7th Advances in Pattern Recognition. Kolkata, India, 2009

23. Siddiqi, I. and Vincent, N. "*A set of chain code based features for writer recognition".* In Proceedings of Document Analysis and Recognition. Barcelona, 2009

24. Hasan, H. Haron, H. and Hashim, S. Z. "*Freeman chain code extraction using differential evolution (de) and particle swarm optimization (pso)".* In Soft Computing and Pattern Recognition. Melaka, Malaysia, 2009

25. Nasien, D. Haron, H. and Yuhaniz, S. S. "*Support vector machine (svm) for english handwritten character recognition*". In Proceedings of Computer Engineering and Application. Bali, Indonesia, 2010

26. Rehman, M. A. U. "*A new scale invariant optimized chain code for nastaliq character representation*". In Proceedings of the 2nd Computer Modeling and Simulation. Sanya, Hainan, 2010

27. Engkamat, A. A. "*Enhancement of Parallel Thinning Algorithm for Handwritten Characters Using Neural Network"*. MSc Thesis, Universiti Teknologi Malaysia, November 2005.

28. Atici, A. A. and Yarman-Vural, F. T. "*A heuristic algorithm for optical character recognition of arabic script"*. Signal Processing*,* 62(1):87-99. 1997

29. Hamid, A. and Haraty, R. "*A neuro-heuristic approach for segmenting handwritten Arabic text".* In Proceedings of Computer Systems and Applications. Beirut, Lebanon, 2001

30. Matthews, P. C, Blessing, L. T. M, and Wallace, K. M. "*The introduction of a design heuristics extraction method*". Advanced Engineering Informatics*,* 16(1):3-19. 2002

31. Yang, C. C and Li, K. W. "*Segmenting chinese unknown words by heuristic method"*. In Proceedings of the 6th Asian Digital Libraries. Kuala Lumpur, Malaysia, 2004

32. Yang, C. C and Li, K. W. "*A heuristic method based on a statistical approach for chinese text segmentation*". American Society for Information Science and Technology, 6(13): 1438-1447. 2005

33. Huang, J. S. and Chuang, K. (1986). *"Heuristic approach to handwritten numeral recognition*". Pattern Recognition, 19(1):15-19. 1986

34. Wegener, I. "*Towards a Theory of The randomized-based Search Heuristics"*. Lecturer Notes in Computer Science, 2747:125-141

35. Hempel, H. "*The randomized-based Algorithms and Complexity Theory*". Computer Science, 12(6):746-761, 2006

36. Sipser, M. "*Time Complexity: Introduction to the Theory of Computation,* 2nd Edition, Course Technology., pp. 1-400 (2005)