A Novel Approach to Angle based Invisible Text Watermarking with EBCDIC Coding

Priyanka Verma
Astt.Professor
MPSTME
SVKM's NMIMS University
Vileparle (W), Mumbai,

Rakhshan Anjum Shaikh
Astt.Professor
MPSTME
SVKM's NMIMS University
Vileparle (W), Mumbai,

Ketki Deshmukh
Astt.Professor
MPSTME
SVKM's NMIMS University
Vileparle (W), Mumbai,

ABSTRACT

Due to excessive proliferation of multimedia, security of digital data has always posed challenges to the researchers. Watermarking of digital data serves the purpose of authentication of ownership and rights of distribution of digital data. A digital watermark is a robust, secure message or a logo, embedded directly into a document; however the watermark may be visible (non blind watermarking) or invisible (blind watermarking). This paper present a new technique of embedding the watermark into the text document where the watermark is mapped as rotation of particular letters in the text based on the proposed lookup table which is prepared using EBCDIC code. The proposed scheme is a blind watermarking scheme which proves to be better in terms of imperceptibility, robustness and fidelity.

Keywords

Digital watermarking, copyright protection, pseudo random number generator, LFSR, EBCDIC coding, Text document.

1. INTRODUCTION

In past years, Internet has become the medium through which thousands of digital data can be easily transmitted, received, duplicated or modified. As a result, copyright protection of digital data is now becomes a legal issue [1].Digital document can be a text, image, audio or video. A lot of work has been carried out for copyright protection of image, audio and video. Although test is the most important part of internet, e-books, articles, legal documents & journals, but its protection has not been given more importance[2].

E-Publishing or digital publishing includes the digital publication of e-books, EPUBs and electronic articles, distribution of written information digitally through CD-Rom's, DVD's, PDF or online over internet or other network. E-Publishing is the creation or distribution and sharing of digital content through a variety of electronic media. The technique came in mind after focusing on above problems is digital watermarking. The basic idea of digital watermarking is to embed digital information into digital data such that it cannot be easily detected and removed. Information hiding is a more general area, encompassing a wider range of problems than the watermarking. The term hiding refers to the process of making the information imperceptible or keeping the existence of information secret. Steganography is an art of concealed communication. In steganography the existence of the message is kept secret.

Digital watermarking emerges out as a solution to protect the copyright of digital documents It is a process of embedding the secret message into original content in order to "mark" its ownership. The secret message embedded into the original content is called the digital watermark. The digital watermark contains data that can be used in various applications, including digital rights management, broadcast monitoring and tamper proofing. The presence of watermark is indicated by the watermark detector when the watermarked content is passed through that. Imperceptibility, Robustness and amount of payload are the three requirements to achieve an efficient and effective watermarking system [3].

The important properties of watermark are-

(i) Robustness of a watermark refers to its ability to survive malicious or non-malicious attacks of removal and common processing tasks like cropping, compression, scaling, resizing etc.

(ii) Data Payload is the encoded message size of a watermark in an image. The simplest form of watermark has no data payload.

(iii) Capacity is the amount of watermark information in the host document. If multiple watermarks are embedded into an image, then the watermarking capacity of image is the sum of all individual watermarks' data payload.

(iv) Imperceptibility is the characteristic of hiding a watermark so that it does not degrade the visual quality of image.

(v) Fidelity is the visual similarity between the watermarked image and the cover image. It depends on the amount of information which is embedded, the strength of watermark and on the characteristics of the original content.

(vi) Security of a watermark is the ability of watermark to resist malicious attacks. These attacks include intentional operation of watermark insertion, modification, removal, and estimation which aim at defeating the purpose of watermarks.

(vii)Computational Cost is the measure of computing resources required to perform watermark embedding and detection processes.

2. TEXT WATERMARKING

Text watermarking is an technique used for text document copyright protection to ensure that the watermark carrying secret information as the copyright so that to prove the ownership of the digital document.[4]. Enormous amount of research work has been contributed to text watermarking [5, 6, 7, 8]. The most commonly used techniques to hide information are line shift coding, word-shift coding, plus feature coding and inter-word space. Text is the most widely used medium of communication existing over the Internet. The major components of websites, books, newspapers, articles, legal documents is simple the plain text. Therefore, plain text requires more protection and security from copyright culprits. In past, a number of digital watermarking algorithms have been proposed for images, audios, and videos; however digital watermarking algorithms for plain text are inadequate and ineffective. The noticeable thing about text watermarking is that the plain text contains less redundant information as compared to images, audio, and video which could be used for secret communication, as happens in steganography, and watermarking. Since plain text is the simplest mode of information, it brings various challenges when it comes to copyright protection. Text has limited capacity for watermark embedding since there is no redundancy in text as can be found in images, audio, and videos. Text document has a binary nature with clear demarcation between foreground and background. The important properties of text are block/line/word patterning, semantics, and structure, style, and language rules. Besides, the inherent properties of a efficient watermarking scheme like imperceptibility, robustness, and security also need to be satisfied.

Brassil, et al. proposed some methods for text watermarking by using text image [5-7]. According to Brassil, the line-shift coding algorithm which alters the document image by moving lines upward or downward (left or right) depending on binary signal (watermark) to be inserted. The second method was about word-shift coding algorithm in which the word moves within text horizontally thus expanding spaces to embed the watermark. The third method is the feature coding algorithm which slightly modifies features such as the pixel of characters, the length of the end lines in characters to encode watermark bits in the text.Maxemchuk, et al. [8-10] analyzed the performance of the methods suggested by Brassil.. The correlation and centroid-based methods [11] are also suggested which treats profiles as a discrete time signal and look for direction of shift and which uses distances between the centroids of adjacent profile blocks for detecting the watermark respectively. Low, et al. [11-12] did the analysis the efficiency of the methods. Young-Won Kim et al. proposed another algorithm based on word classification and inter-word space statistics [13].In this approach, all the words in a text document are classified depending on some text features and then adjacent words comprise a segment and that segment is classified depending on class labels of the words within the segment. Adnan M. Alattar et al. proposed an algorithm [14] to watermark electronic text documents containing justified paragraphs and irregular line spacing.

In our paper we present a novel text watermarking approach based on angle rotation of words and on EBCDIC coding. A pseudo-random watermark is generated by LFSR generator, which is then converted into a secret code by EBCDIC coding.

3. LINEAR FEEDBACK SHIFT REGISTER

LFSR is a pseudo random number generator. An LFSR is a shift register that, when clocked, advances the signal through the register from one bit to the next most-significant bit Fig 1. Some of the outputs are combined in exclusive-OR configuration to form a feedback mechanism. A linear feedback shift register can be formed by performing exclusive-OR on the outputs of two or more of the flip-flops together and feeding those outputs back into the input of one of the flip-flops. Linear feedback shift registers make extremely good pseudorandom pattern generators [15]. A maximal-length LFSR produces the maximum number of PRPG patterns possible and has a pattern count equal to $2^n– 1$, where n is the number of register elements in the LFSR. It produces patterns that have an approximately equal number of 1s and 0s and have an equal number of runs of 1s and 0s.

3.1 8- Bit LFSR Random Number Generator

In this section we discuss about the 8 bit LFSR and its pseudo random generation technique. The most common way to implement a random number generator is LFSR. Codes generated by the LFSR are actually pseudo random sequences because the sequence repeats itself after a certain number of cycles. It is known as the period of the generator. LFSR is based on the recurrence equation,

$$x_{n=} x_{n-1} \oplus x_{n-2} \oplus x_{n-3} \oplus \ldots\ldots\ldots \oplus x_{n-m} \;\ldots\ldots\ldots\ldots eq(1)$$

The operator $\oplus$ is the exclusive OR (XOR) operator. The equation (1) shows that nth bit can be generated utilizing m previous values with XOR operators [20]. The value of m determines the period of the generator. The achievable maximum period is 2^m-1. For the 8-bit LFSR, the recurrence equation is,

$$x_{n=} x_{n-2} \oplus x_{n-3} \oplus x_{n-4} \oplus \ldots\ldots\ldots \oplus x_{n-\delta} \;\ldots\ldots\ldots\ldots eq(2)$$

Since new value x_n depends on previous m values, it is necessary to store previous m values to find the new value.

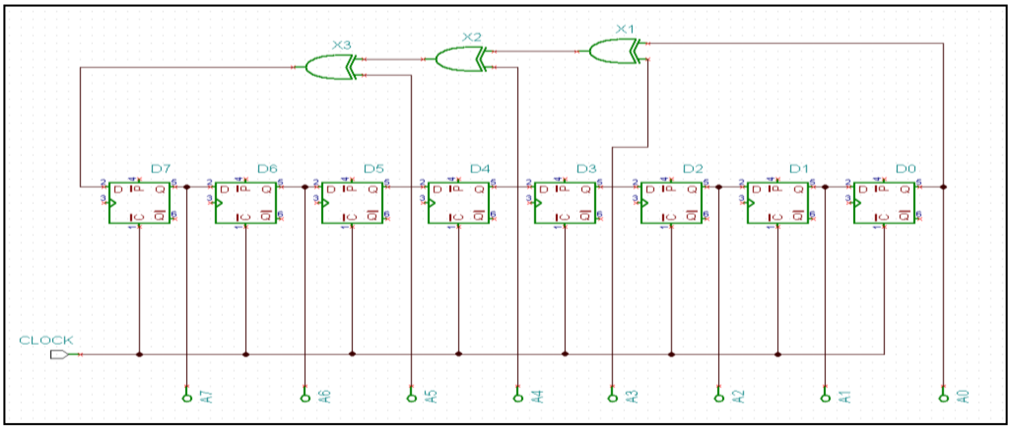


Fig.1 8-Bit Linear Feedback Shift Register

This can be done with m single bit shift registers comprised with flip flops. According to the equation (2), XOR feedback tap positions are taken at 0th, 4th, 5th and 6th flip-flops. The maximum period of the generator is 2^8-1 (255). In each clock pulse, generated new bit is inserted to the shift register while the oldest bit shifts out. Output of the 8 flip-flops form the 8-bit random number. A group of flip-flops connected in series are used with XOR gates to construct the LFSR random number generator (see figure 2). The 8-bit LFSR will generate random numbers which will be coded to give a secret code word. If these random numbers won't match at the extraction process, the original message won't be retrieved [20].

4. LOOK UP TABLE

A modified look up table is created for angle rotation of characters[19], in which we have included all the special symbols also so that number of unique watermark can be generated that will not include the characters only but symbols also. Table 1 shows the look up table in which we number each character after zero decimal because it results in defining the angle of rotation. Zero decimal is used for minor modification in angle and this is result in minor rotation of respective word or character. If we use direct serial number such as 1, 2, 3…. etc, this results in big differences in rotation. Once we get respective position of each character in the look up table, we can use these numbers to insert the watermark into text document[19].

5. WATERMARK GENERATION

The 8 bit LFSR is set by selecting a seed value which will have the values except all 0s, otherwise the LFSR will produce all 0 patterns. After giving the suitable seed value, LFSR is clocked and it will generate a pseudo random sequence of 1s and 0s.The generated pseudo random sequence is then encoded by using EBCDIC code table which will encode the pseudo random sequence into a word. The coded watermark will then embedded into the text document by the proposed approach.

6. PROPOSED METHOD OF TEXT WATERMARKING

Fig.2 shows the proposed flow chart for watermark insertion.

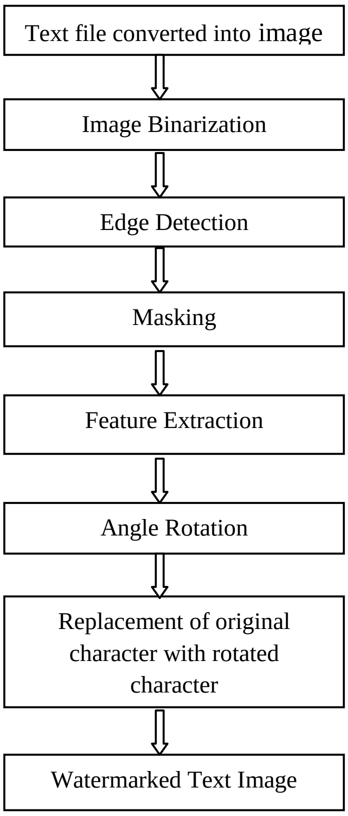
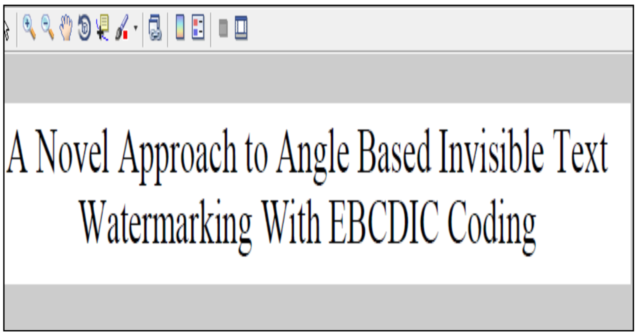
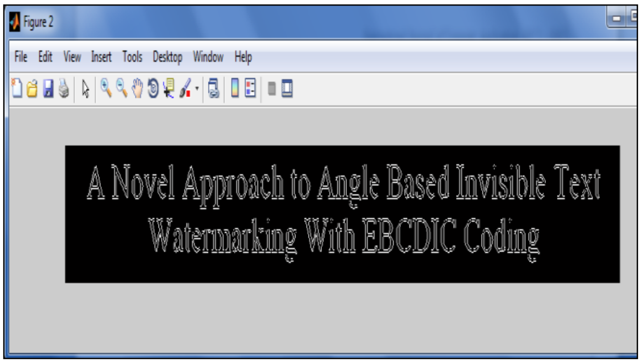


Fig 2.Proposed Flow Chart for watermark insertion



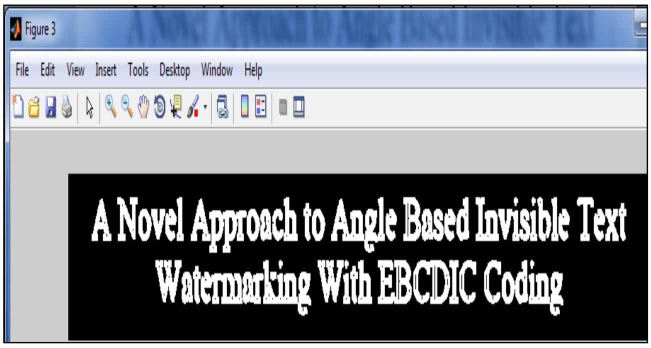
A Novel Approach to Angle Based Invisible Text Watermarking With EBCDIC Coding

Fig 3. Text Document for inserting the watermark



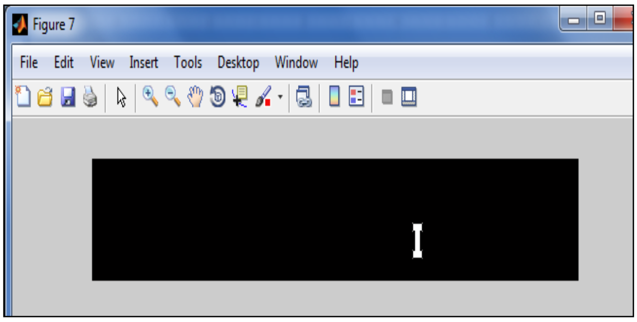
A Novel Approach to Angle Based Invisible Text Watermarking With EBCDIC Coding

Fig.4: Image binarization



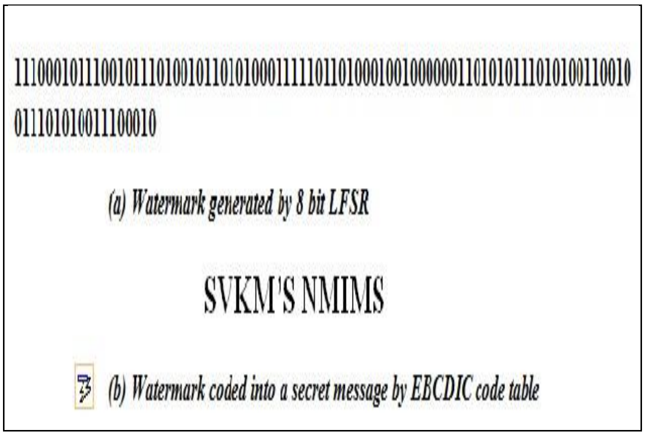
A Novel Approach to Angle Based Invisible Text Watermarking With EBCDIC Coding

Fig.5 Edge Detection



I

Fig 6.Masking performed to get specific character or word



111000101110010111010010110101000111110110100010010000001101010111010100110010
011101010011100010

(a) Watermark generated by 8 bit LFSR

SVKM'S NMIMS

(b) Watermark coded into a secret message by EBCDIC code table

Fig.7 Generation and coding of watermark

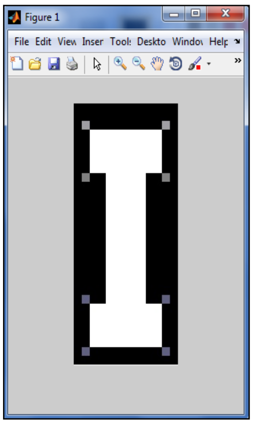


Fig8(a) Extracted Character

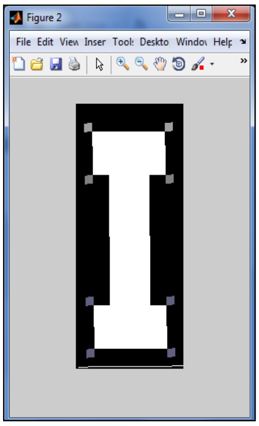


Fig 8(b) rotated by 0.55

7. RESULTS

Fig.3. represents the text document which has to be watermarked.Fig4 shows the text image obtained in the binary form after thresholding.Fig.5 shows the text image obtained after edge detection by applying a suitable edge detector.Fig.6 shows the output when a mask is applied on fig.5 to get the desired character to be rotated.Fig.7 shows the watermark that is generated from 8 bit LFSR and coded using EBCDIC code.

Fig 8(a) the extracted character to be rotated by 0.55 degree since the first character of the watermark is 'S' which corresponds to angle 0.55 degree in lookup table given in Table no.1.Fig 8(b) shows the rotated character which actually represents the watermark insertion.

8. CONCLUSION

In this paper a new technique for text watermarking is proposed which will be faster, reliable and more efficient. It will provide more imperceptibility due to uniform spread of the watermark, and due to rotation angle, which is so minute that is not possible to differentiate between the original one and the watermarked one visually. And also, since the watermark is coded using EBCDIC, it will more robust against channel noise. Due to these two levels processing, the watermark will be more secure and strong.100% fidelity can be achieved by the proposed method as the actual data is not embedded into the main text, only the rotation of words is done. Embedding and detection time is less but the response time is too large for attacker to extract or to modify the watermark.

9. ACKNOWLEDGEMENT

A very special thanks to Prof.Dhirendra Mishra for his support and excellent guidance which really contributed in completion of this research work.

10. REFERENCES

[1] Awanish Kr Kaushik “A Novel Approach for Digital Watermarking of an Image using FFT”,International Journal of Electronics and Computer Science Engineering,ISSN-2277-1956.

[2] Zunera jalil,A.M Mirza, “ An invisible Text watermarking algorithm using image watermark”,Innovations in computer sciences and software Engineering,2010,pp 147-152,print ISBN 978-90-481-9111-6.

[3] C.Patvardhan,A.K Verma,C.Vasantha Lakshmi,”Robust Watermarking of Document and Graphics Images in Wavelet Domain”,IJAIS-ISSN:2249-0868,Vol 2-No.8.June 2012.

[4] ChaoLi Ou, “Text Watermarking for Text Document Copyright Protection”,Computer Science 725 Term paper

[5]. A. Khan and Anwar M. Mirza, *Genetic Perceptual Shaping: Utilizing Cover Image and Conceivable Attack Information Using Genetic Programming*, Information Fusion, vol. 8, no. 4, pp. 354-365, 2007.

[6]. A. Khan, Intelligent Perceptual Shaping of a digital Watermark, PhD Thesis, Faculty of Computer Science and Engineering, GIK Institute, Pakistan, 2006.

[7]. J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O. Gorman, *Electronic Marking and Identification Techniques to Discourage Document Copying*, IEEE Journal on Selected Areas in Communications, vol. 13, no. 8, pp. 1495-1504, 1995.

[8] N. F. Maxemchuk and S. Low, “Marking text documents,” in *Proceedings of the IEEE International Conference on Image Processing, Washington, DC,* Oct. 1997, pp. 13-16.

[9]. N. F. Maxemchuk, S. H. Low, *Performance Comparison of Two Text Marking Methods*, IEEE Journal of Selected Areas in Communications (JSAC), vol. 16, no. 4, pp. 561-572, 1998.

[10].N. F. Maxemchuk, *Electronic Document Distribution*, AT&T Technical Journal, vol. 6, Sept. 1994, pp. 73-80.

[11] S. H. Low, N. F. Maxemchuk, and A. M. Lapone, "Document identification for copyright protection using centroid detection," *IEEE Transactions on Communications,* vol. 46, no.3, March 1998, pp 372-381.

[12].S. H. Low and N. F. Maxemchuk, “Capacity of text marking channel,” *IEEE Signal Processing Letters*, vol. 7, no. 12, pp. 345 -347, Dec. 2000.

[13] Y. Kim, K. Moon, and I. Oh, “A text watermarking algorithm based on word classification and inter-word space statistics”, in *Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDA*R), vol. 2, Aug. 2003, pp. 775-779.

[14] A.M. Alattar, O.M. Alattar, “Watermarking electronic text documents containing justified paragraphs and irregular line spacing”, in *SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, June 2004, pp. 685-694.

[15] W.A.S Wijesinghe,M.K Jayananda and D.U.J Sonnadara ,“*Hardware Implementation of Random Number Generators*,”proceddings of the Technical sessions,22 (2006) 28-38,Institute of Physics-Sri Lanka, pp.28-38

[16] L.Robert,T.Shanmugapriya, *“A Study on Digital Watermarking Techniques*”, International Journal of Recent Trends in Engineering,Vol.1,No.2,May 2009.

[17] Zunera Jalil and Anwar M.Mirza, “ *A Review of Digital Watermarking Techniques for Text Documents”*, IEEE International Conference on Information and multimedia Technology,pp.230-234,2009.

[18] Namrata Rajendra Shah,Aishwarya Vishwakarma, “ *Review on Text Document Watermarking”* ,International Journal of Engineering and Advanced Technology (IJEAT) ISSN:2249-8958,Volume-1,Issue-5,June 2012.

[19] M.M.Lotan, Suryavanshi Hitendra Eknath and Nitin N Patil, *“Angle Based digital watermarking of text Document*”, World Journal of Science and Technology 2012,2(3):171-175 ISSN:2231-2587.

[20] Xiangxue Li,Dong Zheng, and Kefei Chen, “ *LFSR-Based Signatures with Message Recovery,*” International Journal of Network Security,Vol.4,No.3,pp.266-270,May 2007.

Angle	Characters	Angle	Characters	Angle	Characters	Angle	Characters
0.1	0	0.25	o	0.49	M	0.73	(
0.2	1	0.26	p	0.5	N	0.74	)
0.3	2	0.27	q	0.51	O	0.75	_
0.4	3	0.28	r	0.52	P	0.76	\
0.5	4	0.29	s	0.53	Q	0.77	|
0.6	5	0.3	t	0.54	R	0.78	]
0.7	6	0.31	u	0.55	S	0.79	[
0.8	7	0.32	v	0.56	T	0.8	}
0.9	8	0.33	w	0.57	U	0.81	{
0.1	9	0.34	x	0.58	V	0.82	?
0.11	a	0.35	y	0.59	W	0.83	/
0.12	b	0.36	z	0.6	X	0.84	<
0.13	c	0.37	A	0.61	Y	0.85	>
0.14	d	0.38	B	0.62	Z	0.86	;
0.15	e	0.39	C	0.63	&	0.87	:
0.16	f	0.4	D	0.64	-	0.88	+
0.17	g	0.41	E	0.65	~	0.89	=
0.18	h	0.42	F	0.66	!	0.9	
0.19	i	0.43	G	0.67	@	0.91	
0.2	j	0.44	H	0.68	#	0.92	
0.21	k	0.45	I	0.69	$	0.93	
0.22	l	0.46	J	0.7	%	0.94	
0.23	m	0.47	K	0.71	^	0.95	
0.24	n	0.48	L	0.72	*	0.96	

Table.1 Look up Table

LSB▶ MSB ▼	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0100	space										₵	.	<	(	+	|
0101	&										!	$	*	)	;	¬
0110	-	/									¦	,	%	_	>	?
0111										`	:	#	@	‘	=	“
1000		a	b	c	d	e	f	g	h	i						
1001		j	k	l	m	n	o	p	q	r						
1010		~	s	t	u	v	w	x	y	z						
1011																
1100	{	A	B	C	D	E	F	G	H	I						
1101	}	J	K	L	M	N	O	P	Q	R						
1110	\		S	T	U	V	W	X	Y	Z						
1111	0	1	2	3	4	5	6	7	8	9						

Table-2 EBCDIC code Table